NASA CONTRACTOR REPORT

NASA CR-73

DEEP SPACE OPTICAL COMMUNICATIONS SYSTEMS STUDY

Prepared under Contract No. NAS 9-879 *by*
HUGHES AIRCRAFT COMPANY
El Segundo, California
for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • JULY 1964

DEEP SPACE OPTICAL COMMUNICATIONS SYSTEMS STUDY

Prepared under Contract No. NAS 9-879 by

HUGHES AIRCRAFT COMPANY

El Segundo, California

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Office of Technical Services, Department of Commerce, Washington, D.C. 20230 -- Price $3.50

CONTENTS

1. INTRODUCTION

The goal of this initial study phase of the Deep Space Optical Communication Program is to establish the feasibility and desirability of utilizing the laser and suitable optical detectors in a communication system which will provide two-way telemetry, voice, and down-TV from distances of the order of 50 million miles. Since knowledge of both current and extrapolated component and systems capability is required in such a study to suitably define a program for advanced studies, research, and development, several areas of technology had to be surveyed and analyzed. In particular, laser, optical detector, and modulation systems technology had to be carefully explored and boundaries of performance described.

Utilization of the laser with its novel properties required that extension of current communications knowledge be analyzed in terms of supporting equipment. It was found, for example, that the operating frequency, power output capability, spectral purity, etc. required by the system, call for increased performance from optical detectors in regions of the spectrum where no such demand previously existed. Thus, it was necessary to study wide-bandwidth optical detection operating in the infrared wavelengths. Problems in the practical utilization of these detectors in a laser communication system are presented in Section 2. Also presented in this section is a discussion of some of the more complex modulation devices such as the microwave phototube and the crossed-field photomultiplier.

The characterization of the operating properties of the numerous lasers is the subject of Section 3. Not only has the number of lasers increased since the initiation of this study, but so has the number of types of laser. Discussed in this section are pulsed and CW, gas liquid, solid-state and semiconductor lasers with optical, electrical, and other types of pumping mechanisms. An analysis of some of the limitations in bandwidth, spectral purity, and power output is described. The statistical nature of the noise from the various lasers (i.e., its coherence and stability properties) is examined in Section 4.

The laser, with its extremely high frequency provides a potential for communications systems with extremely large bandwidths. This places a requirement for study of the various types of modulation that can be performed and makes necessary careful characterizing of the limitations which

supporting communications technology places on each of these techniques. Section 5 discusses the various types of modulation systems which might be implemented and orders them in terms of their respective virtues. The effect of systems limitations on the ability to fully utilize the enormous bandwidth potential of lasers is indicated in the study of near-range wide-bandwidth modulation given in Appendix 5A.

In Section 6, the properties of various physical devices and their effective utilization in light modulation are discussed. Estimates of the modulator power required, the percent modulation and distortion, and systems limitations are indicated. In Appendix 6A, the limitations imposed by the signal processing circuitry for the GaAs diode used in a PPM mode are discussed in detail.

Section 7 outlines various demodulation systems and compares them theoretically to establish a performance characterization as a function of frequency, bandwidth, and external noise. This receiver performance characterization serves as the basis for the establishment of the optimum receiver selection for various frequency regimes.

Section 8 establishes the characterization of the channel by presenting the effects of attenuation and external noise sources on system performance. Section 9 lists earlier interim reports concerning operations analysis, with emphasis on energy sources, tracking and pointing accuracy, stabilization, and other operational constraints. Section 10 is concerned with the systems analysis and the establishment of the optimum operating frequency selection and the source-destination selection based on results of earlier sections. It is here that the choice of communication link and modulation method is made.

In Section 11, a preliminary systems design of the competitive systems is performed, and comparisons are made. Section 12 consists of summary of the overall results and recommendations for additional work.

2. OPTICAL DETECTOR TECHNOLOGY

INTRODUCTION

The function of the photodetector is to provide an electrical output signal related to the intensity of the incident light signal. In this context the word "light" is used to denote electromagnetic radiation in the visible and the near-infrared range of the spectrum. For the purposes of the contemplated application, the wavelength limits may be set at 0.4 and 4 microns. Over this spectral range several physical mechanisms and devices are available for detection; the utility of any one device is confined to a part of this range. It is therefore necessary to discuss a variety of devices.

The general properties of photodetectors will be discussed, their common basic characteristics identified, and the parameters which are descriptive of their capabilities as detectors or demodulators of light will be defined. Specific detectors will then be discussed in detail.

It is assumed that information is conveyed by an amplitude modulated light beam. The function of the detector is to retrieve the modulating signal. It must therefore act as an integrating device; the period of integration must be long compared to one cycle of the carrier, and short compared to the period of the highest component of the modulating signal. In principle any body which absorbs the incident radiation and which changes some of its physical properties as a result of absorption may act as a detector, but vast practical differences exist among the variety of materials and configurations which may be employed as detectors.

In the infrared region considerable use has been made of thermal detectors. These detectors convert the incident radiation into heat, then the change in the temperature of the detecting element is measured. The merit of such detectors is that they permit an absolute determination of the incident energy; their response is within wide limits independent of the wavelength of the radiation. Other radiation detecting devices are usually calibrated by means of thermal detectors. On the other hand, the integration time of thermal detectors is long; they could only be employed for the demodulation of signals modulated at an impractically slow rate. Because of their slow response and their low sensitivity, thermal detectors will not be considered.

The other group of radiation detectors is based on the photoelectric effect which implies the conversion of the energy of the incident radiation into energy of electrons contained in the detecting element without communicating this energy to the entire material. Photoelectric detectors are quantum detectors because the processes which take place in them preserve the quantum character of the radiation and absorption process. The basic law of every photoelectric effect is that the absorption of energy from the radiation field proceeds in quanta of size $h\nu$, where h is Planck's constant and ν the frequency of the radiation. Moreover, each quantum of energy absorbed from the radiation field is imparted to a single electron.

In the visible part of the spectrum one quantum is generally sufficient to impart enough energy to an electron enabling it to leave a suitable solid. In other words, an external photoelectric effect may take place. The emitted electrons may be collected directly; the current collected is proportional to the intensity of the incident radiation. In the infrared region the energy absorbed from the radiation field is generally insufficient to free the electron entirely, but it may be sufficient to impart to the electron such freedom of motion that it can then be detected by sensing the conductivity of a suitable solid or by sensing the potential difference developed between different points in the solid. Detectors of this type are said to be based on the internal photoelectric effect.

The practice and terminology in the visible and infrared region is somewhat different, but there are common features. In all photoelectric devices operating in their proper range, the voltage or the current output of the device is proportional to the intensity of the input signal. The power output of a photoelectric device operating into a proper load is then proportional to the square of the input radiation intensity. Because of this square law property, photoelectric devices are capable not only of demodulation but of mixing functions as well. However, it must be kept in mind that the output of a photoelectric detector is subject to frequency limitations and that in the case of an extended photodetector the output will be averaged over the variable illumination of the photosurface. Therefore, if two light beams of constant intensity and frequencies ν_1 and ν_2 should be incident on the photodetector, the difference frequency $|\nu_1 - \nu_2|$ will be detected only if the integration time of the photodetector is short compared to $|\nu_1 - \nu_2|^{-1}$ and if the phases of the incident beams do not vary in such an irregular manner over the photosurface that a cancellation of the difference signal will take place upon averaging over the entire surface.

It has already been stated that for light of a fixed wavelength (frequency) the ratio of output voltage (current) to input intensity is a constant. This ratio is called the "responsivity" (R); as a function of wavelength it usually has one maximum. To characterize the spectral variation of the response of a detector, it is customary to provide graphs showing the relative responsivity, i.e., the responsivity for wavelength λ divided by the peak responsivity. The relative response curves are identical for detectors which utilize the identical physical processes and materials, while the peak responsivity may depend on technical characteristics of the instrument as well. When the responsivity

R of an instrument is given without qualification in the visible region, it is understood that this is the value obtainable for the wavelength at which its response is the greatest. Responsivity is expressed in volts per watt or amperes per watt.

There are random fluctuations in every photodetector, which cause the output signal to fluctuate even in the presence of a constant input signal. These fluctuations constitute the noise generated by the detector. This internal noise limits the capability of the detector for the detection of low-level signals. When the noise output is analyzed according to frequency, it is found that for frequencies low compared to the optical ones the noise power output of a detector is proportional to the bandwidth. The noise of the photodetector is characterized by the noise equivalent power (NEP) for one cycle bandwidth. This is the input signal which produces the same output voltage as is present in a one cycle bandwidth due to noise alone.

The NEP is a function of the wavelength of the signal. In the visible range when a figure of NEP is quoted without further specification, it refers to the wavelength for which the responsivity is maximum; in the infrared range it refers to the 500 ° K black-body source. In the former case the NEP can then be computed for another wavelength by means of the relative sensitivity curves according to the equation

$$NEP(\lambda)\, R(\lambda) = NEP\ (peak)\ R\ (peak) \qquad (2\text{-}1)$$

Similar conversion may be made in the infrared for a source of a different temperature or for a monochromatic source. The noise power output of the detector which is proportional to the bandwidth is also proportional to the square of the noise output voltage. The NEP is proportional to the square root of the bandwidth; its proper unit is watt-$sec^{-1/2}$, although frequently the $sec^{-1/2}$ is omitted.

A commonly used figure of merit of a photodetector is the reciprocal of NEP, called "detectivity" D. For most photodetectors the NEP is proportional to the square root of the area; therefore the quantity

$$D^{*} = \frac{A^{1/2}}{NEP} \qquad (2\text{-}2)$$

is a better measure of the detectivity because it is independent of the size of the detector. Modern publications use D^{*}, whose unit is cm-$sec^{-1/2}$/watt.

The physical processes in every detector have a finite lifetime. Together they impose a limitation on the speed with which the output of the detector can be altered by altering its input. Each detector is characterized by a response time, τ, whose reciprocal is the highest angular frequency of

modulation or variation the detector can follow accurately. This is essentially the integration time of the detector. When the modulation frequency is varied, the responsivity varies according to the formula (Reference 2-1)

$$R(f) = \frac{R_o}{(1 + 4\pi^2 f^2 \tau^2)^{1/2}} \qquad (2\text{-}3)$$

where f is the modulation frequency. When $2\pi f = \tau^{-1}$ the responsivity is down to $R_o/\sqrt{2}$, where R_o is its value for low modulation frequencies.

The figures of merit R, D, D^*, and τ are not absolute constants of the material or the physical process. They may depend on the environmental conditions and on the average signal level. For example, the detectivity depends on the temperature of the detector and also on the average output current of the detector. The dependence of the figures of merit on these conditions is different for each kind of photodetector.

The role of the different figures of merit may be summarized as follows: the responsivity curves determine the usefulness of a detector as a function of the propagating field (carrier) wavelength. The peak responsivity determines the output voltage or current level of the instrument. The NEP or detectivity determines the least detectable signal, and the response time τ the highest modulation frequency.

PHOTOMULTIPLIERS

In the visible and near-infrared region of the spectrum the photomultiplier is the most efficient and convenient detector of radiation. It is based on the external photoelectric effect and the subsequent amplification of the electron current by means of a number of secondary emitting stages termed "dynodes." The amplification is simply a matter of convenience: it increases the responsivity of the instrument, but it does not increase its ability to detect weak signals. The basic and determining process in the photomultiplier is the external photoelectric effect, which consists of two steps: absorption of light by a solid and emission of an electron. Light may be absorbed without an electron being emitted; this occurs every time an electron-hole pair is created by absorption of light in a semiconductor.

According to the basic law of the photoelectric effect already mentioned, the emission of each electron is caused by the absorption of a single photon. Electron emission will take place only when the photon possesses sufficient energy to overcome the work function of the solid, i.e., when

$$h\nu \geq q\Phi \qquad (2\text{-}4)$$

where q is the electronic charge and Φ is the work function. This relation sets a long wavelength limit for every photoemissive detector; it may be put in the form:

$$\lambda_L = \frac{12,400}{q\Phi} \tag{2-5}$$

where λ_L is the wavelength limit in angstroms and $e\Phi$ is the work function of the photocathode in electron volts.

The element with the lowest work function is cesium. For this element $q\Phi = 1.9$ ev; therefore, the wavelength limit of a cesium photocathode is about 6500 Å. Composite photocathodes consisting of combinations of metals and oxides have lower work functions; they are capable of functioning up to about $\lambda = 1.2$ microns. Naturally, as the limit is approached the efficiency of the detector decreases. The variation of detector efficiency with wavelength is apparent from the responsivity curves or the curves of quantum efficiency, which in the case of photoemissive detectors are simply related to the responsivity curves.

Quantum efficiency η is the number of emitted electrons divided by the number of incident photons. The frequency ν, quantum efficiency η, incident power P, and the photoelectric current i are related as follows: The number of incident quanta per seconds is $n = P/h\nu$, the electron current emitted from the photosurface is η nq; therefore

$$i = \eta q P/h\nu \tag{2-6}$$

Responsivity $R(\nu)$ is proportional to i/P; therefore

$$R(\nu) = \frac{Kq\eta(\nu)}{h\nu} \tag{2-7}$$

where K is a suitable constant which depends on the amplification of the photomultiplier. For the purposes of this study the quantum efficiency is a more convenient variable than the responsivity. The quantum efficiencies of the red-sensitive photocathodes are shown in Figure 2-1.

The following four wavelengths are particularly important in laser technology: 6328 Å, 6943 Å, 1.06 microns, and 1.13 microns. These are the wavelengths of neon, ruby, neodymium, and neon lasers, respectively. Table 2-1 contains quantum efficiencies and relative responsivities of the photosurfaces S-1, 10, 17, and 20 at these important frequencies.

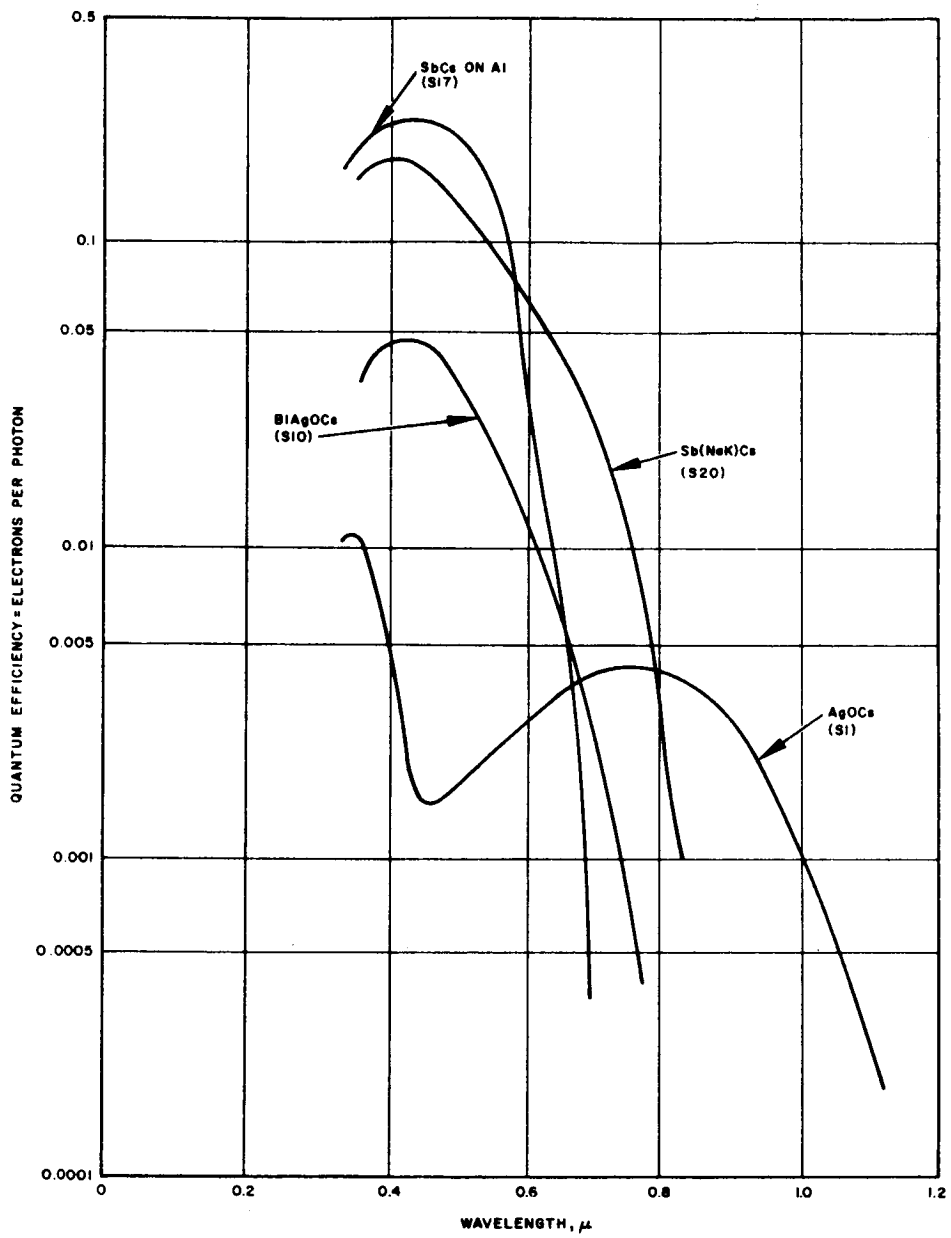


Figure 2-1. Spectral Response of Various Photocathodes

TABLE 2-1. QUANTUM EFFICIENCIES AND RESPONSIVITIES OF RED-SENSITIVE PHOTOSURFACES

	6328 Å		6934 Å		1.06 microns		1.13 microns	
Photo-surface	Quantum Efficiency	Respon-sivity	Quantum Efficiency	Respon-sivity	Quantum Efficiency	Respon-sivity	Quantum Efficiency	Respon-sivity
S-1	0.0033	0.56	0.0040	0.80	0.0009	0.29	0.00016	0.03
S-10	0.0070	0.30	0.0028	0.10	0	0	0	0
S-17	0.0080	0.08	0.0004	0.017	0	0	0	0
S-20	0.045	0.41	0.028	0.22	0	0	0	0

Clearly only the S-1 surface is suitable for 1.06 and 1.13 micron radiation, and the S-17 surface is inferior to the others at the two shorter wavelengths.

A number of photomultiplier tubes are available with these surfaces. The responsivity and the noise equivalent input of the photomultipliers is usually given in microamperes per lumen* and in lumen (Reference 2-2). A test lamp of color temperature 2870° K is employed in place of a monochromatic source to determine responsivity and NEP. The data found in the literature may be converted so that responsivity is expressed in microamperes per watt input and NEP in watt input where the input radiation is that of the peak response. The conversion is accomplished by making use of Engstrom's conversion factors k which take into account the entire spectral response of the photodetectors (Reference 2-3). These factors are listed in Table 2-2.

TABLE 2-2. WAVELENGTHS OF PEAK RESPONSE AND ENGSTROM'S FACTORS

Photosurface	Wavelength of Peak Response	Engstrom's Factor
S-1	8000	93.9
S-10	4500	508.0
S-20	4200	428.0

The data in Table 2-3 are found in Reference 2-2 concerning photomultipliers with S-1, S-10, and S-20 spectral responses:

TABLE 2-3. NOISE IN PHOTOMULTIPLIER TUBES

Tube	Spectral Response	Noise Equivalent Illumination	Temperature, °C
6217	S-10	4.0×10^{-11} lumen	25
7162	S-1	1.7×10^{-12} watt	25
7265	S-20	7.5×10^{-13} lumen	25
7265	S-20	1.0×10^{-13} lumen	-80
7326	S-20	1.9×10^{-12} lumen	25
7326	S-20	3.0×10^{-13} lumen	-80

*One lumen = 1.496×10^{-3} watt.

For tubes of S-10 and S-20 spectral response the figures of Table 2-3 may be converted to noise equivalent power at the peak of the response by means of Engstrom's factors. Then the noise equivalent power for the relevant radiation may be obtained by making use of the data of Table 2-1. In this manner the values of NEP were calculated and are given in Table 2-4.

TABLE 2-4. NEP IN 1-CYCLE BAND AT 6328 Å AND 6943 Å FOR SEVERAL PHOTOMULTIPLIERS

Tube	NEP, watt-sec$^{-1/2}$		Temperature, °C
	6328 Å	6943 Å	
7102	3.0×10^{-12}	2.1×10^{-12}	25
6217	2.6×10^{-13}	7.9×10^{-13}	25
7265	4.3×10^{-15}	8.0×10^{-15}	25
7265	5.6×10^{-16}	10.4×10^{-16}	-80
7326	1.07×10^{-14}	2.0×10^{-14}	25
7326	1.6×10^{-15}	3.2×10^{-15}	-80

For $\lambda = 1.06$ microns the NEP of the 7102 tube is 5.9×10^{-12} watt-sec$^{-1/2}$
For $\lambda = 1.13$ microns it is $5/7 \times 10^{-11}$ watt-sec$^{-1/2}$.

In the course of defining noise equivalent power it has been assumed that the noise originates in an unilluminated photodetector; in short, the dark current noise has been dealt with. Consider now a detector which is already illuminated and ask what noise will interfere with the detection of an incremental signal. In addition to the noise due to the dark current, there will be noise attributable to the random fluctuations of the existing photocurrent. These fluctuations are the consequence of the atomistic nature of electricity: the photocurrent consists of the flow of an integral number of electrons. The rms value of the shot noise current is given by the formula

$$i_s = \sqrt{2qI\Delta f} \qquad (2\text{-}8)$$

where q is the electronic charge (1.60×10^{-19} coul), I is the total current in the photodetector, and Δf is the output bandwidth (Reference 2-4).

The shot current is independent of the nature of the photodetector. It becomes significant when the photodetector is used to detect a small variable illumination in the presence of a constant one (background). Let this constant illumination be P. The corresponding photocurrent I is calculated by means of Equation 2-6. Therefore,

$$i_{sh} = \left(\frac{2q^2 \eta P \Delta f}{h\nu}\right)^{1/2} \tag{2-9}$$

The shot noise equivalent input signal P_{sh} is calculated by making use of Equation 2-6 again

$$P_{sh} = \left(\frac{2Ph\nu\Delta f}{\eta}\right)^{1/2} \tag{2-10}$$

Attention is called to the role of the quantum efficiency η. It does not enter into the calculation of the NEP which pertains to the dark current noise, but it does affect the value of the shot noise.

The absolute responsivity of a photomultiplier depends on the voltages applied to its dynodes. The manufacturers usually state the responsivity under specified conditions in microamperes per lumen, the source of irradiation being a black body at 2870°K temperature. To convert this responsivity into microamperes per watt illumination for a monochromatic radiation it is necessary to divide by Engstrom's k multiplied by the relative responsivity at the wavelength of the monochromatic radiation. The absolute responsivity is of no importance in the detecting capability of the photomultiplier; it does influence, however, the design of the amplifier which follows the photomultiplier. The calculation of the signal-to-noise ratio is based on the NEP of Table 2-4 plus the shot noise calculated from Equation 2-10 plus other noise that may be entering the detector with the signal, including the noise due to signal fluctuations.

The response times of ordinary photomultipliers are between 1 and 3 nanoseconds. From Equation 2-3 it is found that the responsivity is fairly uniform until the frequency $f = 1/2\pi\tau$ is reached. Thus the performance of the commercial photomultipliers begins to be degraded between 50 and 150 mc modulation frequency. Actually, Herriott (Reference 2-5) obtained beat notes up to 300 mc in modulation frequency using an ordinary 7102 multiplier phototube to mix modes of the 1.153-micron neon line. Special tubes are required for the detection or demodulation of signals varying at a faster rate.

MICROWAVE PHOTOTUBES

Siegman and McMurtry (Reference 2-6) have shown that a traveling-wave tube may be used as a phototube. They have obtained outputs up to 4200 mc in modulation frequency when irradiating a Sylvania TW-530 traveling-wave tube with ruby radiation containing several axial modes. Such tubes

could also be used to demodulate amplitude-modulated light with modulation frequency extending into the microwave range.

The traveling-wave phototube is an example of a microwave phototube which consists of a photoemissive surface followed by a microwave amplifier. Such a structure is shown schematically in Figure 2-2. When amplitude-modulated light strikes the cathode it produces a corresponding amplitude-modulated electron beam. This beam is then accelerated and passed through a microwave tube element such as a traveling-wave amplifier. The latter amplifies the modulation of the electron beam and extracts the modulated signal which then emerges at the output. Questions of transit time do not arise in connection with the operation of such a tube because the space-charge wave which represents the modulating signal travels down the tube element where the interaction with the microwave field takes placed.

The cathode of the TW-530 tube is of the BaO:SrO thermionic type; it is not especially suitable for photoelectric emission for incident ruby light. Consequently, it is not surprising that McMurtry and Siegman obtained quantum efficiencies of only 10^{-5} to 10^{-6} with such a tube. The low quantum efficiency is offset by the 40-db gain of the traveling-wave tube.

A photomixer image tube was constructed by Lucy (Reference 2-7). This tube contains an S-1 photosurface which is scanned in the manner of a conventional image tube. The emerging electron beam is directed into a traveling-wave amplifier. Detection of 3.4 Gc beats was reported with a quantum efficiency of 3×10^{-6} at 7000 Å.

An X-band microwave phototube was constructed by Petroff and associates (Reference 2-8). This tube had an S-1 photosurface; the amplification of the traveling-wave tube was 30 db. Signals were obtained at 8.4 and 10.5 Gc. Quantum efficiency data were not published.

There are no published data concerning the noise and responsivity of microwave phototubes. In most instances no suitable amplitude-modulated light input was available for the testing of these tubes, and only semiquantitative data were obtained by means of mixing experiments with light of somewhat uncertain composition.

Dynamic Crossed-Field Photomultiplier

The distribution in electron transit times limits the frequency response of conventional electrostatic photomultipliers to less than a few hundred megacycles. Gaddy and Holshauser (Reference 2-9) have proposed a device in which the electron transit time is rigidly controlled through the use of a microwave electric field to provide the energy for secondary electron emission. The maximum possible bandwidth of their device is half of the electric field frequency.

The high-frequency field is applied across a condenser, one plate of which has been treated to be an efficient secondary emitter as illustrated in Figure 2-3. A photocathode is incorporated at one end of this plate. Electrons emitted when the field is positive will be accelerated toward the

other plate. If a magnetic field B is applied perpendicular to the electric field E, the electrons will be bent into curved trajectories, and for appropriate values of the field E and B and the frequency ω of the electric field some of the electrons will be returned to the first plate and produce secondaries.

The analysis of Gaddy and Holshauser demonstrates that only those electrons emitted at the cathode during the first 60 degrees of a period of the electric field can satisfy the conditions necessary to proceed through a series of multiplications and be collected along with the generated secondaries. The electrons become even more tightly bunched in phase as they proceed through successive steps of multiplication. Signal frequencies in the range $\omega_s < \omega/2$, where ω is the frequency of the microwave electric field, are detected and amplified directly by the device, but frequencies in the range $\omega/2 \leq \omega_s \leq 3\omega/2$ beat with the driving field to produce outputs with frequencies from zero to $\omega/2$. The bandwidth for direct detection and amplification is $\omega/2$.

As with most fast photodetectors developed so far, noise data at high frequencies are not available. Incoherent light modulated at 3 Gc was successfully detected with an experimental device in its mixing mode of operation. A disadvantage of this detector lies in the fact that it samples the photon signal for only one-sixth of each period of the microwave electric field and so must operate at a lower signal-to-noise ratio than a "full-time" electrostatic photomultiplier. Like the other photoemissive devices, it suffers from poor quantum efficiency at the wavelengths of the most useful laser oscillators.

SOLID-STATE DETECTORS UTILIZING INTERNAL PHOTOEFFECTS

Beyond 1.1 microns where photoemissive devices no longer have sufficient sensitivity, detectors utilizing internal photoeffects must be employed. Solid-state detectors with detectivities approaching the theoretical photon noise limit in the wavelength range most likely to be used (0.4 to 4.0 microns) are available as a result of the considerable progress in infrared technology in the past decade. The objectives of this section are to discuss the factors which limit the response time of these detectors, to attempt to evaluate the lower limit to the response time as determined by material properties and the state of art in detector fabrication, and to examine any possible change in detectivity that may result from design for rapid response.

There are two internal photoeffects which are the basis of solid-state detectors operating in the near-intrared range of interest. In both cases, absorption of photons leads to a change in the concentration of free, mobile charge carriers within the material. In the first class of detectors, called "intrinsic," the energy of an absorbed photon creates an electron-hole pair, i.e., the excitation process raises an electron from a valence band state to a conduction band state and only photons with energies greater than the intrinsic band gap are effective. The excitation process in the second class of detectors, called "extrinsic," is the ionization of an impurity center to produce a free carrier and a charged defect site. The optical absorption

constant in intrinsic materials is large, ranging up to 10^5 cm^{-1}, whereas in extrinsic materials it rarely is greater than 10^2 cm^{-1}. Photogenerated carriers are therefore confined to much smaller regions of intrinsic detectors.

Of the extrinsic detectors available, p-type gold-doped germanium is best suited to the near-infrared since its long wavelength cutoff λ_o (defined as the wavelength at which the responsivity decreases to one-half its maximum value) is 7.0 microns. Since in extrinsic detectors majority carriers are generated by the absorbed light, only photoconductive operation is feasible. The time constant of these detectors, set by the recombination time for free carriers with ionized impurity sites, is of the order of 1 microsecond. The low-absorption cross section requires a geometrical configuration such that the capacity of the detectors is of the order of 20 micromicrofarads and the resistance is 1 megohm. Modulation bandwidths greater than 1 megacycle are therefore not attainable with an extrinsic detector without degrading performance severely. It should also be noted that this detector requires refrigeration and does not achieve optimum detectivity unless cooled to 60^o K (below the temperature of liquid nitrogen). The same considerations concerning bandwidth are pertinent to all the extrinsic photoconductors, and those with longer wavelength cutoffs have the additional disadvantage of requiring even lower temperatures of operation.

The intrinsic detectors useful in the near-infrared can be divided into "single crystal" and "thin film" classes for consideration. The latter includes the lead salt series of photoconductors which are available as chemically deposited or evaporated thin films in a variety of geometric configurations. Single crystals of these materials have proved difficult to prepare. The long-wavelength cutoffs of these materials range from 2.5 microns for PbS to 4.5 microns for PbTe. However, their response times are very long, ranging from 10 to 1000 microseconds; the shorter times are available only at the expense of detectivity. The bandwidths required for laser communications are therefore beyond the capability of this group of detectors.

There remain intrinsic single crystal detectors which do offer the possibility of meeting the requirements of a laser receiver system. The long-wavelength cutoffs of detectors that have been fabricated to operate over the spectral range of interest are listed in Table 2-5. Where more than one value appears in the table, each corresponds to a different temperature of operation. Operation of all the detector materials listed depends on the generation of electron-hole pairs when a photon with energy greater than the energy gap between valence and conduction bands is absorbed. Incident radiation can be monitored by "counting" the number of electron-hole pairs generated. The three techniques that have been utilized for the counting process are illustrated in Figure 2-4. Figure 2-4a represents a detector operated in the photoconductive (PC) mode. The dashed line represents the steady-state electron-hole concentration in a uniform block of material. The current that flows in response to an electric field applied between the

TABLE 2-5. LONG-WAVELENGTH CUTOFF FOR VARIOUS DETECTOR MATERIALS

Detector Material	Long-Wavelength Cutoff, λ_o*
Si	0.8
GaAs	0.9
GaSb	1.5
Ge	1.8
InAs	3.7
InSb	7.3 (300° K) 5.6 (77° K)

* λ_o is the wavelength at which reponsivity has decayed to one-half its peak value.

electrodes provides a measure of the free carrier concentration. With signal photons incident on the detector, an additional concentration of carriers is set up varying with position in the detector as shown by the solid curve in the figure. Because of the additional carriers a larger current flows in the external circuit.

The photoelectromagnetic (PEM) mode of operation is obtained if the electric field is removed from the photoconductor so that it is connected directly across a load and a magnetic field is applied perpendicular to the plane of the figure. Because of the concentration gradient of electron-hole pairs produced by the absorbed radiation in the material, carriers drift in the direction in which radiation is incident. The Lorentz force due to the magnetic field on the moving carriers deflects the electrons and holes to opposite electrodes and produces current in the external load, as illustrated in Figure 2-4b.

The third mode of operation is the photovoltaic mode in which a p-n junction is produced immediately behind the surface on which radiation is incident by diffusing a p-type dopant into n-type material (Figure 2-4c). The generated electron-hole pairs diffuse to the junction under the influence of the concentration gradients set up in response to the incident radiation. The electric fields in the junction drive the electrons to the n-side and holes to the p-side of the detector. With zero applied bias an external photocurrent can be observed or the detector may be operated with reverse bias in what is essentially a photoconductive mode.

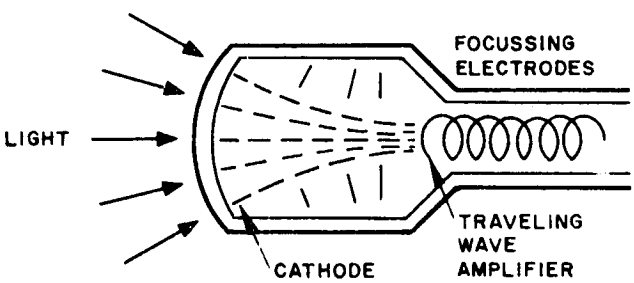


Figure 2-2. Microwave Phototube

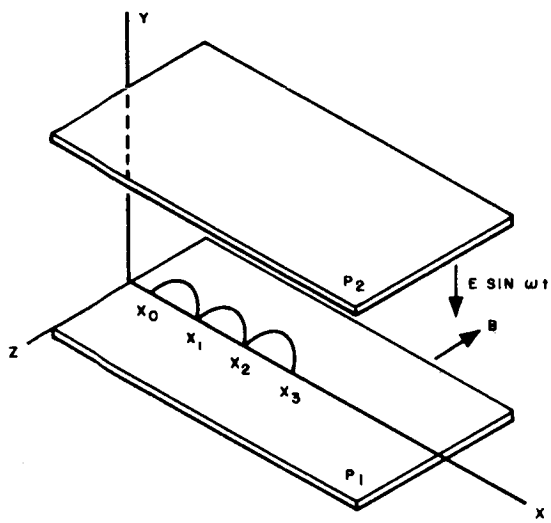
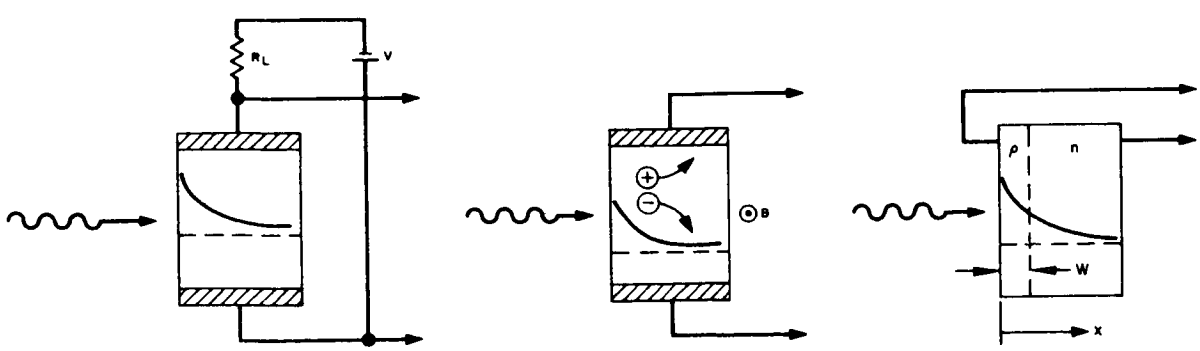


Figure 2-3. Physical Configuration of Electron Multiplication System



a) Photoconductive mode b) Photoelectromagnetic mode c) Photovoltaic mode

Figure 2-4. Three Modes of Operation of Solid-State Photo Detectors

By appropriate control of the fabrication process, it is possible in some semiconductors to make the transition region between the n and p regions large or even to arrange that the detector consist of three regions: a thick n-type base region on which there is first a moderately thick intrinsic region and then a very thin p region at the surface on which radiation is incident. Such a p-i-n structure has somewhat different response time characteristics when biased in the reverse direction and is particularly useful for weakly absorbed radiation.

For a more detailed analysis and review of the properties of infrared detectors see References 2-1 or 2-9.

FACTORS DETERMINING RESPONSE TIME OF PHOTODETECTOR

The response time of a detector will depend on the mean time between the generation of excess carriers by light and their collection at the electrodes in PC and PEM cells. The requirement for fast response and maximum responsivity determines the geometry of the detectors. To achieve the short response times required for laser communications applications requires very small structures, and since these have been achieved most successfully in photovoltaic detectors, the preceding remarks will be illustrated by a more detailed consideration of such devices.

Consider first the p-n junction cell illustrated in Figure 2-4c. A photon flux density F_o is incident on the front surface of the cell giving rise to an excess concentration of generated electron-hole pairs inside the cell that varies with position according to the relation

$$n = F_o (1 - \rho) \alpha e^{-\alpha x} \tag{2-11}$$

where ρ is the reflection coefficient of the semiconductor surface and α is the absorption constant for the wavelength of interest. The excess carriers will diffuse in the concentration gradient created, and the resulting particle flux density will be given by the diffusion equation

$$\Psi_n = -D_n \frac{dn}{dx} \tag{2-12}$$

D_n is the diffusion constant of the excess minority carriers and in the case of the p-n junction device illustrated it is the diffusion constant of electrons in the p-type layer adjacent to the surface. It is related to the electron mobility by the Einstein relation $D_n = \mu_n kT/q$ (where k is Boltzmann's constant, T the absolute temperature, and q the magnitude of electronic change). Since particle flux density can also be written as the product of concentration and velocity, $\Psi_n = nv$, the velocity is found to be

$$v = -\frac{Dn}{n}\frac{dn}{dx} = +\frac{kT}{q}\mu_n \alpha \qquad (2\text{-}13)$$

The response time of the detector may be estimated as the time it would take for a carrier generated at the surface to diffuse to the junction:

$$\tau = \frac{\beta}{v} = \frac{q\beta}{kT\mu_n \alpha} \qquad (2\text{-}14)$$

where β is the depth of the junction below the irradiated surface. For a germanium detector at room temperature looking at the 1.15-micron He-Ne gas laser line, the parameters have the following values

$$\alpha \sim 10^4 \text{ cm}^{-1}$$

$$\mu_n = 3900 \text{ cm}^2/\text{volt-sec}$$

$$T = 300^\circ \text{ K}$$

whence it is found that $\tau \sim 10^{-9}$ second for a junction depth β of 10 microns. It has been tacitly assumed in the above analysis that the diffusion length of minority carriers in the material is large compared to ω, or, stated in another way, the mean lifetime against recombination is larger than the transit time τ. If this were not the case, some of the photo-generated carriers would recombine before reaching the junction and would not contribute to the electrical output of the detector. The electron-hole recombination time τ_o is a function of the impurity content of the material and varies from a few milliseconds to 10^{-8} second in most semiconductors. For materials used in detectors it can generally be assumed that $\tau_o > \tau$ and the corresponding diffusion lengths are larger than the junction depths.

In the example cited the absorption constant was so large ($\alpha = 10^4$ cm^{-1}) that most of the incident radiation was absorbed within a few microns of the surface. If technologically possible the junction could be brought to within 2 microns of the surface with no appreciable loss in signal but at a considerable improvement in response: τ would be reduced by a factor of 5 to 2 x 10^{-10} second. Alternately, a laser line at a wavelength such that α were only 10^3 cm^{-1} could be detected as effectively by the original detector with a response time of 10^{-9} second.

The p-I-n form of the photovoltaic detector presents the possibility of more efficient utilization of the radiation for a more weakly absorbed line. Thus, suppose the I-region extends from 2 microns below the surface to 10 microns below the surface. When a large reverse bias is applied across the detector, essentially the complete voltage drop appears across the intrinsic region and a line at wavelength such that $\alpha = 10^3$ cm^{-1} generates carriers throughout the intrinsic region (Figure 2-5). Because the I-region is very

narrow, the electric field intensity for several volts reverse bias will be of the order of 10^3 volts/cm or larger. In such fields the average drift velocity of carriers in a semiconductor increases and eventually reaches a saturation velocity, v_o, which in germanium is 6×10^6 cm/sec (References 2-10 through 2-13). Estimating the response time of the detector as the time required for a photo-generated carrier to drift across the junction gives $\tau \sim 1.3 \times 10^{-10}$ second.

It may not be possible to realize the transit time response of the photovoltaic detector because of resistor-capacitor (RC) limitations encountered in the circuit in which it is incorporated. A reverse-biased junction has a capacity determined by the area and width of the junction and the dielectric constant of the material:

$$C = \frac{8.88 \times 10^{-12} \epsilon/\epsilon_o A}{W} \quad \text{farad} \qquad (2\text{-}15)$$

where

ϵ = dielectric constant

A = area of junction

W = width of junction in meter-kilogram-second units

The p-I-n junction used for illustrative purposes above would have a capacity of 1.75×10^3 pf/cm^2. Diffused junctions with areas of 2×10^{-4} cm^2 can be fabricated, at least in germanium (Reference 2-14), and such a junction would have a capacity of 0.35 pico-farad. The series resistance of the n-type base region in such a detector is in the range of 1 to 10 ohms so that the characteristic RC time of the device is 3.5×10^{-12} second. This will be degraded if larger load resistors are coupled to the detector.

More detailed analyses of the operation of photovoltaic detectors are available in the literature. A good review of various modes of photodetector operation can be found in Reference 2-1. The p-n junction detector is covered in detail by D. E. Sawyer and R. H. Rediker (Reference 2-15). The basic work on the p-I-n structure was reported by W. W. Gartner (Reference 2-16). The graded junction diode, which may be regarded as a cross between the above two, is described by A. G. Jordan and A. G. Milnes (Reference 2-17). On the basis of such analyses and the current state of the art with respect to material preparation and device fabrication, Lucovsky, Lasser, and Emmons (Reference 2-18) have estimated the 3-db cutoff frequencies of a variety of photovoltaic detectors for available laser lines ranging from the 6943 Å ruby line to the 2.49 microns U^{+3}:CaF_2 line. Their estimates place the frequency cutoffs between 4.6×10^9 cps and 7.3×10^{10} cps. They consider Si, Ge, and GaAs operated at room temperature, InAs at 196° K, and InSb at 77° K.

Detailed reports of operating detectors are rather sparse. The most complete description of such a detector published (Reference 2-14) describes a Ge p-I-n junction with a 2.5-inch thick I-region 0.006 inch in diameter. This detector exhibited a pulse rise time of 0.6 nanosecond and detected beats up to 2 kmc in the output of a pulsed ruby laser.

A point contact p-π-n diode (Reference 2-19) with a π region 5 microns thick and 5 microns in diameter has been used to detect beats up to 936 mc in a He-Ne laser operating at 1.15 microns (π indicates a lightly doped p region).

Saito et al (Reference 2-20) have used a standard Ge parametric diode to detect 4-kmc beats from ruby, and 11-kms beats from such a source have been detected with a silicon p-I-n diode (Reference 2-21). Lucovsky, Lasser, and Emmons (Reference 2-18) have reported beat measurements to 1.46 kmc with Ge, Si, GaAs, and InAs photodiodes.

NOISE CONSIDERATIONS IN FAST PHOTOCELLS

Equivalent Circuit of Reverse-Biased Photovoltaic Detector

As noted earlier, the photovoltaic mode of operation will generally be preferred for very fast detectors because of the relative ease of fabricating structures of the required small dimensions. The current voltage characteristic of such a detector has been found to be

$$I = I_s \left[\exp \frac{qV}{kT} - 1\right] - I_{sc} + \frac{V}{R_{sh}} \qquad (2\text{-}16)$$

The first term on the right is the I-V characteristic of a p-n junction in the dark; the second term is the additional reverse saturation current due to optically generated carriers; and the third term represents the current through the shunt conductance of the diode. When operating into the small loads required to achieve wide-band performance, the small series resistance r_s of the base region must also be taken into account and the constant-current equivalent circuit of such a detector in the reverse-biased region with a load R_L is as shown in Figure 2-6. r_p is the parallel resultant of the dynamic resistance,

$$\left.\frac{\partial V}{\partial I}\right]_F ,$$

and the shunt resistance of the diode; C is the diode capacity. The constant-current generator has a magnitude:

$$i_s = \frac{\partial I}{\partial F}\Delta F = -\frac{\partial I_{sc}}{\partial F}\Delta F \qquad (2\text{-}17)$$

where F is the incident photon flux density. The inductance L is added to the circuit so that the diode capacitance can be tuned out for operation at higher frequencies.

The number of carriers created per second is ηAF, where η is an overall quantum efficiency, A is the area of the detector, and the current produced by these carriers is $I_{sc} = \eta qAF$ (q is the magnitude of the electronic charge). The short-circuit current is $i_s = \eta qA\Delta F$.

Reverse-bias operation is preferred to photovoltaic operation when wide bandwidths are sought because r_p is higher and C is smaller than in the latter case.

If the parallel combination of r_p and C are lumped together in an impedance $Z_p = r_p (1 + j\omega C r_p)^{-1}$, and similarly the load resistor and inductance L in parallel are treated as an impedance $Z_L = R_L (1 - j R_L/\omega L)^{-1}$, then the rms signal voltage into the amplifier for ∂ photon signal at frequency ω of rms value ΔF is

$$N_s = \eta Aq\Delta F \frac{Z_p(r_s + R_L)}{Z_p + r_s + R_L} \frac{R_L}{r_s + R_L} \qquad (2\text{-}18)$$

When r_s is small compared to R_L, which is generally the case, the signal bandwidth, defined as the frequency range in which $v_s > 0.707\, v_{s,max}$, is given by $\Delta f = (2\pi R_L C)^{-1}$.

Noise Equivalent Circuit of Semiconductor Diode

Three types of noise must be taken into account in treating photodiodes.

1) At low frequencies current noise, having a power spectrum given by

$$\overline{i_I^2} = \frac{KI^2}{f}\Delta f \qquad (2\text{-}19)$$

usually predominates. However, for most photodiodes fabricated with sufficient care this noise is negligible above a few kilocycles and so may generally be disregarded in discussing wide-band laser receivers.

2) The most important source of noise in a photovoltaic detector is shot noise caused by the particle nature of the current. It has a flat (up to frequencies determined by transit time effects) power spectrum given by

$$\overline{i_{sh}^2} = 2q\bar{I}\Delta f \tag{2-20}$$

where $\bar{I}$ is, as before, the average diode current and q is the electronic change.

3) The third source of noise which must be considered is thermal noise in the various resistive elements in the diode circuit. Since current noise can be neglected at the anticipated high frequencies of operation, the noise equivalent circuit of a photodiode can be drawn as in Figure 2-7 with various current generators having the values:

$$\overline{i_d^2} = \left[2qI + \frac{4kT_d}{r_p}\right]\Delta f$$

$$\overline{i_s^2} = \frac{4kT_d}{r_s}\Delta f$$

$$\overline{i_L^2} = \frac{4kT}{R_L}\Delta f \tag{2-21}$$

The voltage generator $\overline{e_n^2} = 4kTR_n\,\Delta f$ represents the amplifier noise in terms of an equivalent input resistance R_n. T_d is the temperature of the detector, T that of the load.

The rms equivalent input noise voltage to the amplifier is given by

$$\frac{|e_n|^2}{\Delta f} = \left[2qI + \frac{4kT_d}{r_p}\right]\frac{|Z_p|^2\,|Z_L|^2}{|\Sigma|^2} + 4kT_d\,r_s\,\frac{|Z_L|^2}{|\Sigma|^2}$$

$$+ \frac{4kT}{R_L}\,\frac{|Z_L|^2\,|r_s + Z_p|^2}{|\Sigma|^2} + 4kTR_n$$

$$\Sigma = r_s + Z_L + Z_p$$

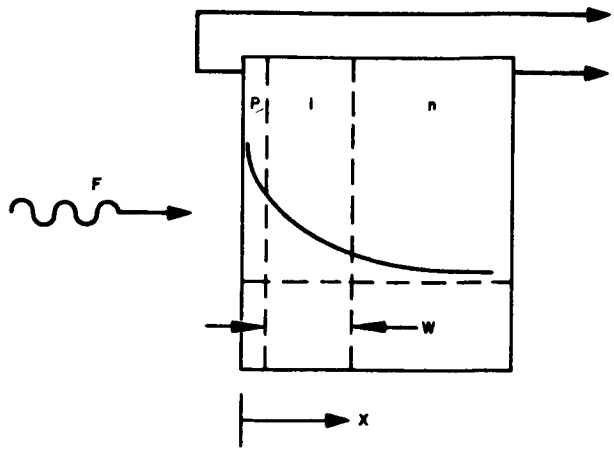


Figure 2-5. P-I-N Junction Detector

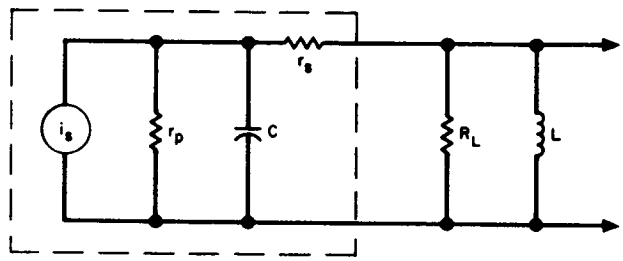


Figure 2-6. Constant-Current Signal Equivalent Circuit of Photodiode

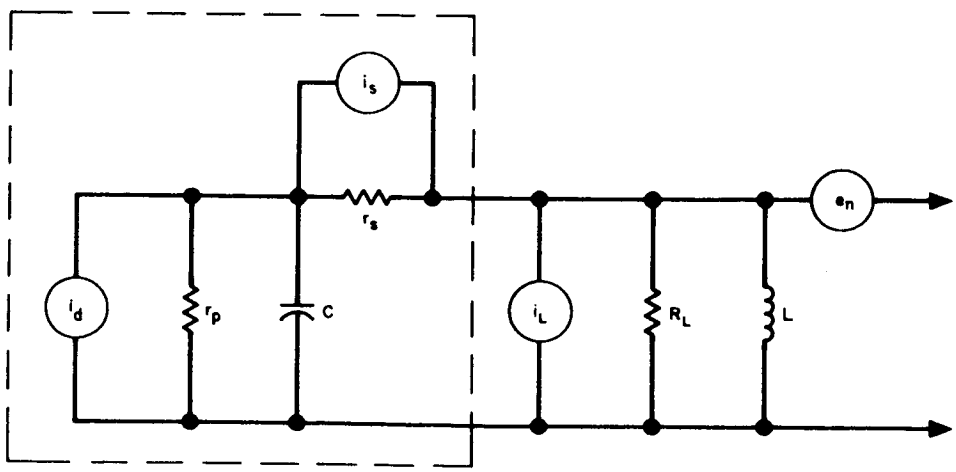


Figure 2-7. Noise Equivalent Circuit for Photodiode

By equating $|e_n|^2$ to $|v_s|^2$, an expression for the minimum detectable photon flux can be found:

$$\phi_{min}^2 = A^2 \Delta F^2 = \Delta f (\eta q)^{-2} \left\{ 2qI + \frac{4kT_d}{r_p} + 4kT_d r_s \frac{1}{|Z_p|^2} + \frac{4kT}{R_L} \frac{|r_s + Z_p|^2}{|Z_p|^2} + 4kT R_n \frac{|\Sigma|^2}{|Z_p|^2 |Z_L|^2} \right.$$

A considerable simplification results if the idealized case where r_p approaches infinity and r_s goes to zero is considered.

$$\phi_{min}^2 = \Delta f (\eta q)^{-2} \left\{ 2qI + \frac{4kT}{R_L} \left(1 + \frac{R_n}{R_L} \right) + 4kTR_n \left(\omega C - \frac{1}{\omega L} \right)^2 \right\}$$

In the vicinity of the resonance frequency, $\omega_o = (LC)^{-1}$, the principal sources of noise are 1) shot noise caused by diode current, 2) thermal noise in the load resistor, and 3) additional noises introduced by the amplifier. An interesting question to consider is: What diode current is necessary so that shot noise will be at least as large as the thermal noise in R_L? Assuming a 50-ohm load as fairly typical of microwave circuits, it is found that $I_{min} = 4kT/2qR_L \simeq 10^{-3}$ ampere. If the amplifier makes a significant contribution to the noise, I_{min} is increased by the factor $(1 + R_n/R_L)$. The photon flux necessary to produce this current is $AF = I/\eta q = 1.3 \times 10^{16}$ photons/sec for a quantum efficiency of one-half. This corresponds to an incident power of 2.3×10^{-3} watt at a wavelength of 1.15 microns and 7.7×10^{-4} watt at 3.39 microns.

On the other hand, thermal noise in the load leads to a minimum detectable photon flux

$$\Phi_{min} = \frac{1}{\eta q} \sqrt{\frac{4kT}{R_L}} = 2.3 \times 10^8 \text{ photons/sec}$$

for a 1-cps bandwidth. For a 100-mc bandwidth, the minimum detectable power is 4×10^{-7} watt at 1.15 microns and 1.4×10^{-7} watt at 3.39 microns. The load has been assumed to be at room temperature for these calculations.

The conclusion from the above is that photodiode performance as an envelope detector will always be limited by thermal noise in the load and amplifier. This limitation can be overcome in heterodyne operation if a

local oscillator with the power outputs noted above is available. As shown by Forrester (Reference 2-22) and Oliver (Reference 2-23), both the shot noise power and the signal power increase in the same proportion as the local oscillator power is increased. It is therefore possible to achieve a condition where shot noise is dominant. The price that has to be paid for this improvement in performance is in the requirement for close alignment of signal beam and local oscillator beam. Constructive interference between the two beams can occur only if they are aligned within an angle $\Delta\theta \leq \alpha/d$, where λ is the wavelength of the light and d is the diameter of the collecting optics.

As an indication of what could be achieved with semiconductor photodiodes, the best currently available information concerning dynamic resistance and D* for the materials most likely to be used in the visible and near-infrared has been added to Table 2-6. A word of caution is necessary in using these numbers. The data reported are from standard infrared detector evaluation procedures and so are with load resistors that approximately match the dynamic impedance. In broad-band envelope detection, thermal noise in the smaller loads required may severely degrade detectivity, as pointed out above. In heterodyne operation, the listed detectivities may well be achieved in practice.

Also included in Table 2-6 is an estimate of the longest wavelength λ_c at which a particular material may be used to produce fast photodetectors. This limit is chosen, on the basis of transit time considerations, as that wavelength for which the optical absorption constant is 1000 cm^{-1} at the operating temperature of the diode.

To assure optimum electrical characteristics, the material with its λ_c closest to (but greater than) the laser wavelength being used should be chosen. Fabrication of such a detector in an appropriate package may still not be a trivial problem, and in no case have experimental performance limits (response time, detectivity in wide-band operation) yet been reported.

SUMMARY

P-n or p-I-n junction devices can probably achieve the highest speeds of various modes of operation of solid-state photodetectors, because the technology for producing structures of the requisite small dimensions is currently available. These devices are especially useful at wavelengths greater than 1.1 microns where photoemissive devices are no longer operative and quantum efficiencies approaching one can be attained with semiconductor detectors. Disadvantages to be considered in their application are their capacity, which restrict the bandwidth over which they can be operated, and their small sensitive area. In addition, heterodyne operation is required to attain optimum sensitivity, and this imposes severe restrictions on the alignment of the signal radiation with respect to the local oscillator. Detectors operating at longer wavelengths (2 to 4 microns) also require cooling for maximum detectivity.

Laser System	Emission Line	Gallium Arsenide 300° K	Silicon 300° K	Germanium 300° K	Indium Arsenide 196° K	Indium Antimonide 77° K
Ion:Host	Microns	← Structure Critical Dimension 3 db Frequency →				
$Cr^{+3}:Al_2O_2$ $Sm^{+3}:CaF_2$	0.6943 0.7083	P-N $W_p = 2 \times 10^{-5}$ cm $\omega = 4.7 \times 10^{10}$	P-I-N $W_I = 3 \times 10^{-4}$ cm $\omega = 2.9 \times 10^{10}$	P-N $W_p = 2.5 \times 10^{-4}$ cm $\omega = 3.1 \times 10^{11}$	P-N $W_p = 10^{-4}$ cm $\omega = 1.1 \times 10^{11}$	P-N $W_p = 10^{-4}$ cm $\omega = 4.6 \times 10^{11}$
$Nd^{+3}:CaF_2$ $Nd^{+3}:CaWO_4$ $Nd^{+3}:SrMoO_4$ Ne:He	1.046 1.063 1.064 1.15			P-I-N $W_I = 10^{-4}$ $\omega = 1.4 \times 10^{11}$	P-N $W_p = 10^{-4}$ $\omega = 1.1 \times 10^{11}$	P-N $W_p = 10^{-4}$ $\omega = 4.6 \times 10^{11}$
$Tm^{+3}:CaWO_4$ $Ho^{+3}:CaWO_4$	1.91 2.046				P-N $W_p = 10^{-4}$ $\omega = 1.1 \times 10^{11}$	P-N $W_p = 10^{-4}$ $\omega = 4.6 \times 10^{11}$
$U^{+3}:CaF_2$	2.49				P-N $W_p = 10^{-4}$ $\omega = 1.1 \times 10^{11}$	P-N $W_p = 10^{-4}$ $\omega = 4.6 \times 10^{11}$
λ_c ($\alpha = 10^3$ cm^{-1}), microns	-	0.88	0.80	1.53	3.2	4.2
D^*_{peak}	-	2.9×10^{11}	3.4×10^{13}	10^{10}	5×10^9	1.5×10^{10}
NEP ($A = 10^{-4}$ cm^2), watts	-	3.5×10^{-14}	3.0×10^{-16}	10^{-12}	2×10^{-12}	6.7×10^{-13}

NOTES:

1) Four upper rows taken from Reference 2-18. Detectivity data as reported in various Photodetector Reports, Naval Ordnance Laboratory, Corona. D* for Ge dates from 1955 and is probably not representative of best available today.

2) W_p = depth of function behind the irradiated surface; W_I = width of the intrinsic region.

TABLE 2-6. DYNAMIC RESISTANCE AND DETECTIVITIES OF

REFERENCES

2-1. Kruse, McGlauchlin, and McQuistan, Infrared Technology, John Wiley and Sons, 1962.

2-2. RCA Electron Tube Handbook.

2-3. R. W. Engstrom, "Absolute Spectral Response Characteristics of Photosensitive Devices," RCA Rev. 21, 184 (1960).

2-4. D. K. C. MacDonald, Phil. Mag. 40, 561 (1949).

2-5. D. R. Herriott, "Optical Properties of a Continuous Helium-Neon Optical Maser," J. Opt. Soc. Amer. 52, 31 (1962).

2-6. B. J. McMurtry and A. E. Siegman, "Photomixing Experiments with a Ruby Optical Maser, etc.," Applied Optics 1, 51 (1962).

2-7. R. F. Lucy, "An Experimental Photomixer Image Tube," Proc. IEEE 51, 162 (1963).

2-8. M. D. Petroff, H. A. Spetzler, and E. K. Bjornerud, "X-Band Microwave Phototube for Demodulation of Laser Beams," Proc. IEEE 51, 614 (1963).

2-9. O. L. Gaddy and D. F. Holshouser, Proc. IEEE 51, 153 (1963).

2-10. "IRIA State of the Art Report on Infrared Quantum Detectors," edited by W. Wolfe and T. Limperis, published by Inst. of Science and Technology, Univ. of Michigan, Ann Arbor, Michigan (Contract NOnr 1224(12)) (July 1961). (Classified Confidential)

2-11. E. J. Ryder, Phys. Rev. 90, 766 (1953).

2-12. J. B. Gunn, J. Electron. 2, 87 (1956).

2-13. J. B. Arthur, A. F. Gibson, J. W. Granville, J. Electron. 2, 145 (1956).

2-14. R. P. Riesz, Rev. Sci. Inst. 33, 994 (1962).

2-15. D. E. Sawyer and R. H. Radiker, Proc. IRE, 46, 1122 (1958).

2-16. W. W. Gartner, Phys. Rev., 116, 84 (1959).

2-17. A. G. Jordan and A. G. Milnes, IRE Trans. on Electron Dev., ED-7, 242 (1960).

2-18. G. Lucovsky, M. E. Lasser, and R. B. Emmons, Proc. IRE, 51, 166 (1963).

2-19. L. U. Kibler, Proc. IRE, 50, 1834 (1962).

2-20. S. Saito, K. Kurokawa, Y. Fuju, T. Kimura, Y. Uno, Proc. IRE 50, 2369 (1962).

2-21. H. Inaba, A. E. Siegman, Proc. IRE 50, 1823 (1962).

2-22. A. T. Forrester, J. Opt. Soc. Am. 51, 253 (1961).

2-23. B. M. Oliver, Proc. IRE 49, 1960 (1961).

3. LASER TECHNOLOGY

INTRODUCTION

With the development of laser oscillators, it is now possible to generate coherent radiation at a large number of optical frequencies at considerable power levels. Laser oscillations have been obtained in many different types of materials. The most frequently used materials have been of the solid state and gaseous type. Laser oscillations have also been obtained in glasses, optical fibers, liquids, organic materials, and semiconductors.

This section discusses the general characteristics and major types of laser oscillators.

LASER CHARACTERISTICS

Coherence and Spectral Properties of Lasers*

Optical laser beams are thought to be suitable for applications to communications due to the high degree of temporal and spatial coherence exhibited by the laser emission. Since the discovery of lasers, the properties of the laser emission have been studied extensively. Some present knowledge of the coherence and spectral properties as well as the detailed structure of the laser emission that may be of special interest in applications to optical communications will be discussed.

Usually many modes oscillate simultaneously in an optical resonator. These modes have different frequencies and different spatial characteristics so that the use of more than one mode in a specific application may not be possible. When all the modes except one are suppressed, the power that was formerly distributed over many modes is concentrated in a single mode. This mode suppression has been accomplished in a He-Ne gas laser by Kogelnik and Patel (Reference 3-1). They have suppressed the unwanted longitudinal modes without decreasing the gain of the cavity so that the output power is

*See Section 4, Statistical Study of Laser Transmitters for a more detailed discussion of spectral and stability characteristics.

not decreased and is concentrated in a single frequency near the center of the doppler-broadened line. Theoretically, when the oscillator output consists of a single longitudinal mode, the light is spatially coherent across the beam cross section and the collimation of the beam is thus limited by diffraction. The beam spread of the emission of a He-Ne gas laser with 4 milliwatts continuous power in the 1.153 micron line has been observed to be almost diffraction limited (Reference 3-2). The line shape consists of three or more components less than a few hundred cycles in width and separated by the spacing of modes in the interferometer.

The spatial coherence of a beam can be measured by observing the far field or multiple slit diffraction patterns while the time coherence is measured by the spectral line width. The mutual correlation function between any two points in the beam is given by a combination of these two effects. However, the line width of the emission is not always indicative of the time coherence since the output may consist of unresolved off-axis and axial modes. These off-axis modes have been observed by Ready (Reference 3-3) and many others. In the ruby laser one or more sharp components have been observed (Reference 3-4) with a spectral width of 6×10^{-4} A. McMurtry and Siegman (Reference 3-5) have made observations on the output of a ruby laser that indicate that the line width is about 2 mc. The resonant cavity, formed by the end faces of the ruby, is able to support many resonant frequencies within the atomic line breadth. Adjacent modes are separated by about 602 mc in the laser used in this experiment. Photo mixing of the output produced microwave signals spaced about 600 mc apart showing that the laser is oscillating in several simultaneous axial modes separated by the proper frequency interval.

Interference patterns have been formed by the direct illumination of two slits with the light emission from the end face of a ruby. The formation of interference fringes demonstrates the phase coherence between different points on the face of the ruby (Reference 3-6). Kisliuk and Walsh have made beams from the opposite ends of a ruby laser interfere and have obtained interference fringes (Reference 3-7). This implies a constant phase relationship between points at different ends of the ruby, the expected result for a system of standing waves in a cavity. The output of a ruby laser consists of a series of spikes of about a microsecond duration separated by a microsecond or longer intervals. The distribution of light across the face of the ruby, known to be nonuniform, need not be the same from spike to spike. If the distribution of the light does change, the interference pattern would change from spike to spike and observations made over a large number of spikes would not show interference fringes. Hence, the existence of fringes in the experiments of Kisliuk and Walsh imply, in addition to constant phase relationship, that the distribution of light across the face of the ruby must be nearly the same for all spikes.

Irregularities of the fringe patterns are also observed, which is consistent with the results of Nelson and Collins who found that the fringe patterns were displaced. Both these observations are accounted for by assuming a slight phase difference between spatially separated points on the ruby face.

The spectral purity of the emission of gas lasers is much greater than that of the ruby output. Javan et. al. (Reference 3-8) have found the intrinsic line width of a He-Ne gas laser to be not larger than 2 cps. The theoretical limit of the narrowness of the spectral lines is a consequence of a certain unavoidable amount of spontaneous emission. The measured line width, even though extremely small, is still several orders of magnitude greater than the theoretical limit. It should be emphasized, however, that this greater degree of monochromaticity of the gas laser does not mean that gas lasers necessarily possess a clear advantage over pulsed lasers for the purposes of communication. Present modulation techniques cannot make use of a fraction of the potential bandwidth available at optical frequencies, and spectral purity is not the only factor to be considered in a communications system.

The filamentary nature of the emission of ruby lasers has been frequently observed. High speed photographs (Reference 3-9) of the emission patterns have been made which show that light is emitted in a pattern of small spots on the ruby face, and that the emission pattern does not change much during the pumping pulse. This is in agreement with deductions made from the observation of interference fringes discussed above. In addition, streak photographs of sufficient resolution to show the structure of a single laser spike were made. They show that within each spike is a regular oscillation of about 50 mc, and that each emitting area of the ruby produces a spike with the same 50 mc oscillation on it, with the oscillations sometimes out of phase.

Power and Operating Characteristics

Optical masers are presently limited to either high-power pulsed operation or relatively low-power continuous operation, and the power efficiencies of these coherent optical frequency sources, with the exception of the Ga As lasers, are still very low (less than 1 percent). What is needed is a moderately efficient, medium-power continuous source, or perhaps a high-power, efficient, pulsed source with a high-duty cycle. There is presently a considerable effort towards developing sources with these capabilities; Hughes Research Laboratory at Malibu, California and R. C. A. Laboratory at Princeton, New Jersey, are presently concerned with the development of 10 watt continuous solid-state lasers of improved efficiency. The achievement of high-duty cycles requires flash tubes capable of repetitive high-power operation and the maintenance of low temperatures in the laser crystal. Since the power dissipated in the crystal is of the order of the power output, and efficient operation of the laser requires low temperatures, cooling the crystal is not a minor problem. The problems of repetitive pulsing will not be discussed, but some features of the single-pulse operation of pulsed lasers and, in particular, the giant pulse mode of operation will be considered.

In the usual pulsed operation of a ruby laser and other solid-state lasers, the output consists of a series of microsecond duration spikes spaced several microseconds apart. The individual spikes have a peak power of the order of a few kilowatts and the total energy in the series of spikes is about

a joule. The spikes or oscillations continue so long as the input power is above the threshold for oscillation. Usually the ruby is pumped by flash lamps in pulses about a millisecond long.

In normal laser operation, oscillation occurs when the gain of the cavity exceeds the losses by a small amount. The stimulated emission causes an increased radiation density in the cavity and a subsequent depletion of the population of the excited metastable state. The population of the excited state is then decreased more rapidly than it is replenished by the pump photons. When the population of excited states is reduced below the critical number required to provide net gain in the cavity, the oscillation subsides. This is a rough description of the production of a single laser spike. Continued pumping repopulates the metastable state and the process is repeated as long as the pumping energy is sufficiently high to obtain a net gain in the cavity. The spiking is not periodic even when the pumping energy is essentially constant. That is, there are variations in the amplitude and spacing of the laser spikes within a single pulse. There have been numerous explanations for the spiking phenomena, none of which seem to completely resolve the problem. It was originally thought that the spiking was a damped oscillation and would decay when the ruby was operated continuously. This has proved not to be the case, since in the CW operation of a ruby laser the spikes have been observed to persist with essentially the same characteristics that are observed in pulsed operation. Both damped spiking and continued undamped spiking have been observed in other solid-state lasers that have operated CW. ($Ca F_2$: Dy^{2+}, $Ca WO_4$: Nd^{3+}, $Ca F_2$: U^{3+})

The output of a giant pulse ruby laser consists typically of a single high-power short-duration burst of coherent radiation. These pulses are about 10^{-8} seconds or longer, and peak powers of the order of megawatts are easily obtained when proper switching techniques are used. A single pulse occurs only for special switching conditions on the power regeneration to the cavity, and the output often consists of a large pulse followed by several low-power pulses. The giant pulse mode of operation is achieved by varying the power regeneration to the cavity, which is equivalent to varying the Q of the resonating cavity.

To obtain giant pulses, a ruby laser contained in an optical cavity of very high loss is pumped. This can be arranged by making the reflection losses of the cavity very large. A large population inversion in the ruby to obtain net gain and laser oscillation in the cavity is then required. When a large population inversion is obtained and before oscillation takes place, the reflection losses are decreased, producing a situation where the normal threshold conditions for oscillation are greatly exceeded. When the switching of the power regeneration is sufficiently fast, a single short-duration high-power pulse is produced. The detailed structure of the pulse is governed by some fairly complicated laser dynamics and line broadening effects that are not fully understood. However, many of the gross features of the pulse structure are understood. It is known that the pulse shape is sensitive to the cavity Q switching time and that for sufficiently slow switching several low-power pulses follow the initial intense pulse. McClung and Hellwarth

(Reference 3-10) have investigated the effect of the switching rate on the pulse power as a function of time, and they have also studied the effect of the switching rate on the spectrum of the emitted radiation. The pulses of shortest duration have the broadest spectrum. The spectral width of the R_1 ruby line of the giant pulse is broader than the line width for normal laser operation, and the beam width of the giant pulse is essentially the same as for normal lasers, 1.3., about a milliradian.

In principle the energy output per pulse per unit volume of laser material should be about $1/2\ h\nu\ (\mu_2-\mu_1)$ where $h\nu$ is the energy per photon and $\mu_2-\mu_1$ is the population inversion density. For pink ruby and an initial inversion of 50 percent, about one joule output per cubic centimeter of ruby is expected. The observed output of the giant pulses are far less than the expected results. Apart from the discrepancies between theoretical and experimental energy emission, there are basic limitations in the energy per pulse that can be obtained from a single ruby laser. In principle it is impossible to achieve a high initial inversion over arbitrarily large dimensions, limiting the size of the ruby that may be used effectively. For example, if a very long ruby is used, a high inversion will be achieved in the central portion and not in the ends. Hence, only the middle of the ruby will participate in the laser action. The depumping of the ends is caused by spontaneous emission from the ends undergoing a large gain as it passes the length of the ruby. When the ruby is long enough the inversion at the ends is depleted in a manner independent of the reflection from the end faces of the ruby or cavity.

The threshold for laser operation is determined by the minimum gain necessary to overcome the loss mechanisms existing in a given cavity. In turn, the gain is directly related to the energy or power extracted from the source. Hence, the threshold is measured in joules or watts, according to the mode of operation.

Since devices vary widely in size and configuration it is difficult to make meaningful comparisons between reported threshold values. Doping, pump source, coupling, reflectivity, crystal size, operating temperature, and other factors all have a considerable effect on the threshold. Generally, it increases with doping, crystal size, and operating temperature, while it decreases with increased reflectivity.

There are a number of different efficiencies that can be defined for lasers, but generally the one cited is the net conversion efficiency (source input power to laser output). However, there are two factors which set a limit on the net conversion efficiency and these factors are characteristic of the laser medium.

Quantum efficiency is the probability that an atom in one of the pump band states will make the transition to the terminal state via the laser transition. In general there is a finite possibility that the atom will make the transition by reradiating the pumping radiation. The other limiting mechanism is the energetic efficiency which is defined as the ration of the laser transition photon energy to the pumping transition photon energy (for optical pumping).

Generally, the quantum efficiency is ~ 70 percent while the energetic efficiency may be as low as 10 percent. The product of these two sets a theoretical limit on the net conversion efficiency of any device that may be constructed.

The doping or concentration (number/cm^3 or percent by weight) determines the maximum energy which can be stored in a unit volume since each atom can store an amount of energy $h\nu$. The energy storage capacity can be increased by increasing the doping, but at the same time, the energy required to obtain equal populations will also increase, resulting in higher thresholds. Also, the power required for CW operation will increase since the power loss per unit volume due to spontaneous decay is directly proportional to the number of atoms in the excited state. For these reasons, it has been found advantageous to use dopings of less than 1 percent.

Properties of CW Lasers

Laser action on a continuous or quasi-continuous basis has been demonstrated in various types of active media such as solid state dielectrics, semiconductors, and gases. Some of the presently available media are listed in Table 3-1, along with other properties basic to the laser action itself, without regard to particular devices. It is seen that a wide variety of frequencies, conversion efficiencies, and pumping characteristics is available.

Characteristics of particular devices which have been successfully operated are given in Table 3-2.

SOLID-STATE LASERS

Properties of Solid-State Lasers

Solid-state lasers are usually excited by broad-band absorption of optical frequency radiation. A fast nonradiative transition then populates sharp fluorescence levels. When a sufficient population of excited states is obtained, stimulated emission takes place with the subsequent emission of coherent narrow beam radiation.

When the terminal level of the radiative transition is the ground state, the optical maser is called a three-level system. Ruby is an example of a three-level laser. In four-level lasers, the terminal level lies above the ground state. Almost all solid-state lasers are of the four-level type. To obtain maser oscillation in a three-level system, a slight inversion of the excited and ground states is necessary, and to maintain the inversion, at least as much power must be supplied to the cavity as is lost by spontaneous decay of the upper level and the nonradiative transition. The power emitted by spontaneous decay is proportional to the number of excited states and, since in a three-level system the number of excited states required for oscillation is more than half the number of dope-ions, it takes a large input power to maintain the oscillation.

TABLE 3-1. CHARACTERISTICS OF CONTINUOUS WAVE LASERS

Medium	Pumping Mechanism	Emission Frequency, microns	Efficiency	
			Energetic, percent	Quantum, percent
$CaF_2:u^{3+}$	0.88 to 0.92 micron	2.613	35	–
$CaF_2:Dy^{2+}$	0.23 to 0.49 micron (strongest) 0.58 micron 0.72 micron 0.91 micron	2.36	10 20 30 38	–
$CaWO_4:Nd^{3+}$	0.59 micron	1.065	55	–
$Al_20_3:C_R^{3+}$	0.38 to 0.43 micron 0.5 to 0.6 micron	0.694	57 80	~70
G As:(Zn, Te)	Electron injection	0.84	–	~100
He-Ne	Gas discharge	1.15 0.63	–	–

TABLE 3-2. OPERATING CHARACTERISTICS OF CONTINUOUS WAVE LASERS

Medium	Threshold	Power		Frequency, microns	Temperature, °K
		Input	Output		
$CaF_2:u^{3+}$	2000 watts	2 kilowatts	10 microwatts	2.613	77
CaF_2Dy^{2+}	100 watts	1 kilowatt	350 milliwatts	2.36	27
$CaWO_4:Nd^{3+}$	1200 watts	1.5 kilowatts	500 milliwatts	1.065	85
$Al_2O_3:C_R^{3+}$	850 watts	930 watts	4 milliwatts	0.694	77
GaAs	50 milliwatts	50 milliwatts	25 milliwatts	0.84	4.2
He-Ne	50 watts	50 watts	15 milliwatts	1.15	–

In a four-level system when the terminal level is far enough above the ground state, its population will be negligible. It then requires a correspondingly smaller population in the excited state to obtain maser oscillation. The power lost to spontaneous decay will then be less with the consequence that the threshold for oscillation can be relatively small. Other factors such as narrow absorption band, low quantum efficiency, and scattering of radiation also influence the threshold for maser oscillation.

Pulse Operation

The power output of solid-state lasers is usually obtained in the form of pulses. In this type of operation, the laser is usually irradiated for a short period of time with a broad-band optical source. In the experiments with ruby performed by Maiman, et al (Reference 3-11), the excitation light from a flash lamp is appreciable for times of the order of millisecond. The stimulated emission lasts for some fraction of that time. The emitted pulse of radiation is not smooth; it consists of spikes or oscillations occurring with a frequency that depends on the instantaneous intensity of the flash tube light and doping density. The amplitude of the spikes is somewhat erratic, and the average frequency of the spikes is about 2×10^6 cps and the length of the whole pulse is about 0.5 millisecond. Hence there are about two or three hundred oscillations in a single pulse. The largest peak power obtained in the experiments was 5 kilowatts with an energy of nearly a joule in the pulse.

Divalent and trivalent rare earths in suitable host materials generally satisfy the characteristics necessary for maser action. Trivalent uranium in several different hosts have also been made to laser. Table 3-3 is a partial summary of solid-state materials that have produced laser emission and a tabulation of some of the characteristics of their operation. The pulsed operation of the four-level systems is essentially the same as that described for ruby. The four-level masers tend to be more sensitive to temperature, since to obtain efficient operation, the terminal level often has to be depopulated by cooling. As in ruby, the output consists of spikes in the main pulse. These oscillations often appear strongly damped in contrast to ruby where no damping has been observed. In one case ($Ca F_2$: Dy^{++}) there appears to have been critical damping of the oscillations, as they were not detected (Reference 3-12). Usually the frequency of the oscillations is fairly constant with some fluctuations of the amplitude. A pulse that consists of very regular damped oscillations has been observed (Reference 3-13) in $Ca WO_4$: Nd^{++}, but because of the narrow absorption band this laser does not seem suited for high power operation. Threshold energy for pulse operation is generally much lower than the 800 joule threshold in ruby. At low temperatures, threshold energies of the order of a few joules have been observed in several different maser systems (Table 3-1).

The threshold energy is sensitive to changes in temperature. The dependence of threshold for maser action in $Ca F_2$: U^{3+} has been measured by Boyd, et al (Reference 3-14). The flash lamp energy required to obtain maser oscillation was: 20 joules at 20°K, 3.78 joules at 77°K, 4.35 joules at

TABLE 3-3. PROPERTIES OF SOLID-STATE LASERS

Material	Type, level	Wavelength, microns	Threshold Energy to Flash Pump, joules	Temperature, °K	Remarks	Reference
$Al_2: Cr^{3+}$	3	0.6943	695	77	Continuous operation 77°K, 850 watts flash power threshold	3-10, 3-8
$CaF_2: U^{3+}$	4	2.613	2.0 3.78 4.35 1200	20 77 90 300	Continuous operation 77°K, 1200 watts threshold flash power	3-4
$CaF_2: Sm^{2+}$	4	0.708	5	20		
$CaF_2: Dy^{2+}$	4	2.36	0.1 1	4 77	Continuous operation 27°K, 0.5 watt output, 50 watts threshold flash power	3-2, 3-9
$CaF_2: Tm^{2+}$	3 4	1.116 1.189	50 50	4.2 4.2		3-22
$CaF_2: Nd^{3+}$	4	1.046	80	77		3-7
$SrF_2: U^{3+}$	4	2.407	8 32 38	~20 77 90	No maser oscillation at 300°K with 2000 joules flash energy	3-23

TABLE 3-3 (continued)

Material	Type, level	Wavelength, microns	Threshold Energy to Flash Pump, joules	Temperature, °K	Remarks	Reference
$CaWO_4: Nd^{3+}$	4	1.063	3 5	77 300	Continuous operation ~ 85°K, 1300 watts flash power	3-3
$CaWO_4: Pr^{3+}$	4	1.0468	70	77		3-2
$CaWO_4: Tm^{3+}$	4	1.911	125	77		3-24
$CaWO_4: Ho^{3+}$	4	2.046	300	77		3-2
$CaWO_4: Er^{3+}$	3	1.612	800	77		3-2
$SrMoO_4: Nd^{3+}$	4	1.064	42 125	77 300		3-25
$SrMoO_4: Pr^{3+}$	4	1.047	~ 70	77		
$LaF_3: Nd^{3+}$	4 4	1.063 1.0399	93 75	77 77		

90° K, and about 300° K. Temperature dependence of the threshold energy has also been observed for most of the other materials.

The nonradiative transition results in absorption of energy by the crystal lattice, and hence in continuous high power operation there will be a large amount of energy transferred to the crystal lattice. Maintaining constant temperature under these conditions presents a difficult cooling problem. Heat transfer from the maser material will probably be of practical importance in attaining high power repetitive pulsing in solid state materials.

Giant Pulse Operation

In normal pulsed operation, oscillations begin as soon as a critical inversion is achieved, that is, as soon as the single pass gain exceeds the single pass loss. Oscillation may be postponed by making reflection losses very large and if the ruby is strongly excited a high inversion can be established. A Kerr cell switching technique has been used to decrease the cavity loss suddenly. Consequently, there is inversion far above threshold with the consequence that the energy in the cavity is dumped in a very short high power pulse. McClung and Hellwarth (Reference 3-15) have obtained peak output powers of 600 kilowatts with a pulse duration of 12 microseconds. More recently, Woodbury (Reference 3-16) using a slightly different arrangement has measured peak powers of 20 megawatts of pulse duration 7×10^{-9} seconds.

In principle, one can control the length and power in a single pulse of this type. The power out would be roughly inversely proportional to the pulse duration. There is a limit to the peak power obtainable, which is set by the maximum energy that can be stored in the ruby and the minimum time it takes to dump it. This time cannot be less than the time it takes a photon to traverse the length of the cavity several times.

Ruby is especially well suited to the giant pulse mode of operation because of the relatively long lifetime of the fluorescence level.

Continuous Operation

In continuous operation of four-level solid-state materials, the oscillations that are characteristic of the pulsed mode are damped out in times less than a few milliseconds, or they have not been observed at all because of inadequate time resolution of the detecting device. This is in contrast to ruby where the oscillations persist with a fairly regular frequency and a large slightly erratic amplitude. The characteristics of the oscillations are similar to those observed in the pulse mode of operation.

Continuous power output has been obtained with several different maser materials. The system $Ca F_2 : U^{3+}$ has operated continuously at 77°K (Reference 3-14) with output power of 10 microwatts. Threshold for continuous oscillation is 1200 watts and maser action occurs at a wavelength of 2.613.

Trivalent neodymium in host material Ca WO_4 has lasered continuously at 77°K with 1300 watts flash tube power (Reference 3-17) and output of 1 milliwatt at 1.065 microns. The system Ca F_2: Dy^{++} has also been made to oscillate continuously (Reference 3-12). In the first reported operation, the output was about 2 milliwatts at 2.45 microns. At 27°K the threshold for continuous operation was 600 watts electrical input to the lamp.

Continuous operation has also been attained in a ruby optical maser (Reference 3-18) with the maser oscillation at 0.6943 micron. The threshold for oscillation at 77°K was 850 watts supplied to a short arc high-pressure mercury lamp. At an input power 950 watts, the maser beam produced 4 milliwatts of power, and as mentioned above consisted of undamped oscillations.

It has been reported recently that threshold for continuous operation with Ca F_2: Dy^{++} has been reduced to 50 watts using a tungsten lamp as the pump (Reference 3-19). Using a 1 kilowatt tungsten lamp, continuous power of about 1/2 watt has been obtained. The power is concentrated in a single mode when operated at liquid Ne temperature and power efficiency with respect to energy absorbed by the material of about 1 percent. Considerably higher powers are expected with improved cooling and optimized crystal geometry. It is clear that Ca F_2: Dy^{++} is presently one of the most promising materials for high power continuous output.

Ruby is another material that at least theoretically should be capable of high continuous power output (Reference 3-20). The development of a 10-watt continuous ruby laser has been proposed, and theoretical considerations indicate that this should be possible without any radically new technological developments. Efficiencies of about 1 percent are anticipated. So far, relatively little effort has gone into obtaining high efficiency, high continuous power, or high power repetitive pulsing. As a result, perhaps, efficiencies reported in the literature are less than 1 percent are usually less than 0.1 percent. There does not appear to be a basic limitation that would prevent substantial increases in efficiency in the normal course of the development of laser technology. Townes has speculated that efficiencies of 10 or 20 percent are possible. (Reference 3-21).

Related to the problem of efficiency is the production of high power, or more precisely high continuous or quasi-continuous power. As discussed above, it is presently possible to obtain very high power pulses of short duration. However, the rate at which these pulses may be repeated is presently limited by the power capabilities of the excitation source and by heating of the ruby and optical elements in the system. The attainment of high power repetitive pulses or continuous power output depends, among other things, on the development of higher efficiencies, adequate cooling techniques, and improved excitation sources.

Spectral Characteristics

Optical maser oscillations have been obtained in the range 0.6943 to 2.613 microns. The output exhibits spacial and time coherence with beam angles that are typically about 10^{-2} radians. There is presently not much information in the literature about the line width and frequency stability of the output of solid-state materials with the exception of ruby.

In ruby the spectral width of the light output appears to be set by the 0.5 microsecond interval between spikes as the measured bandwidth of the emission from a single spike is about 2 mc.

When the spectral distribution of the radiation from a number of spikes was measured, a bandwidth of 20 mc was obtained, which showed that there was no coherence between successive pikes. It has been observed that the ruby oscillates simultaneously in a number of axial modes (Reference 3-22). There is evidence of spacial coherence over the end face of the ruby of the order of 100 wavelengths (Reference 3-23).

The frequency of the ruby line is temperature dependent which indicates thermal tuning is possible. Measurements of the temperature dependence of the frequency of the R_1 ruby line show that the change of wavelength with temperature is 0.065 Å/deg (Reference 3-24). Because the spectral width is of this order, good temperature control is necessary for frequency stability.

Ruby Lasers

The normal emission from a ruby laser typically consists of a pulse of duration of the order of a millisecond, the pulse duration being established by the length of time a flash tube emits sufficient pumping intensity to maintain the oscillation in the cavity. This millisecond pulse is made up of narrow spikes that occur, with what appears to be a somewhat random frequency and amplitude. The spikes are often about a microsecond duration and successive spikes are usually separated by several microseconds or more. For most methods of modulation this signal is not suitable as a carrier in an optical communication channel. The envelope of the emission is not predictable and secondly, the presently possible duty cycles are too low to give adequate average powers.

The causes of the spiking are beginning to be understood. Recently a regular damped oscillation has been obtained by a mode selection technique, which will be discussed later. Further, a truly periodic oscillation has been obtained in a confocal ruby rod; more precisely, the spiking is periodic during the short interval that the pump intensity is constant. Presumably, if these techniques can be employed in continuous operation, a constant amplitude emission would be obtained in the first case and a periodic oscillation in the second. Both of these types of emission would provide a suitable carrier signal. However, there has been only one report of CW ruby operation and this emission has been low power (several milliwatts) and displays the

undamped irregular spiking of the type usually observed in the pulse operation of ruby. The difficulty in obtaining CW operation is connected with the high threshold pump power intensity required in three-level lasers. Nevertheless ruby still has many properties that make it attractive for high average power operation and there has been considerable optimism on the attainment of CW power of from 1 to 10 watts.

One of the essential problems associated with CW operation of ruby, as well as other solid materials including glass and semiconductors, is the transfer of energy dissipated in the material during oscillation. In solid-state lasers, for output powers of a watt, more than a watt will be dissipated in the crystal and therefore the crystal must be cooled to continue maser action. The cooling of the crystal and the development of excitation sources capable of maintaining the high intensity pump light necessary for maser oscillation are considered to be the major problems in the development of high-power, solid-state CW lasers.

Observations of Damped and Periodic Spiking

The undamped oscillation of the laser emission from ruby and solid materials in general is now thought to be connected with the fact that oscillation can occur simultaneously in many modes. Single mode operation has been obtained (References 3-25 and 3-26) in a traveling-wave maser of the type shown in Figure 3-1. Experimental results showed that when the optical isolator was removed from the system the usual irregular spiking occurred. When the optical isolator is used to suppress unwanted modes, a regular damped oscillation is obtained. Off-axis modes are not suppressed by the isolator and it is believed that irregularities in the emission observed at high pumping power are caused by the excitation of these modes.

A periodic output has been obtained using ruby rods with confocal reflectors (Figure 3-2). No damping of the spikes was detectable within the duration of the pumping time. Beam divergence for confocal resonators are greater than for Fabry-Perot type, the divergence depending on the radius of curvature of the end focus relative to the crystal length. This behavior is quite different than that observed by Johnson in experiments with N_d 3+ $CaWO_4$, where spiking was suppressed when confocal geometry and high pump powers were employed.

Recently laser operation has been obtained in a ruby ring (References 3-27 and 3-28). The emission consists of a regular damped oscillation similar to that obtained by Tang et al (Reference 3-25) with the traveling-wave laser except that damping occurs much faster in the ring geometry than in the traveling-wave laser. The difference in damping can be explained by assuming a very high Q for the ring laser; this assumption is consistent with the small losses that occur with total internal reflection.

The line width of ruby operated in the normal pulsed manner has been found to be 2 megacycles which corresponded to the laser spikes in this experiment of 0.5 microsecond duration (Reference 3-29). If the line width

is set by the width of the spikes a further narrowing of the stimulated emission line when the spiking is damped out is expected. Unfortunately no experimental data on the spectral and directionality properties of damped oscillatory ruby emission exist in the published literature.

Dynamics of Laser Emission

The dynamical equations, first proposed by Statz and de Mars (Reference 3-30), that have been partially successful in describing the transient and time-independent operation of rubies and other solid and gaseous lasers will be considered. The equations governing the laser system are:

$$\frac{d\rho}{dt} = c\sigma\left(\mu_2 - \frac{g_2}{g_1}\mu_1\right)\rho - \frac{\rho}{t_L} \tag{3-1}$$

$$\frac{d\mu_2}{dt} = -c\sigma\left(\mu_2 - \frac{g_2}{g_1}\mu_1\right)\rho + W\mu_o - A\mu_2 \tag{3-2}$$

where ρ is the photon density of the coherent radiation in the cavity. The population densities of the upper and terminal laser levels and the ground state are given by μ_2, μ_1, and μ_o respectively. Losses in the cavity are accounted for by the decay time t_L, the increase in photon density in the cavity and depopulation of the excited level μ_2 due to stimulated emission is given by the first terms on the right-hand side of Equations 3-1 and 3-2 respectively, where σ is the cross section for stimulated emission, and g_1 and g_2 are factors that account for degeneracy in the levels 2 and 1. The probability of spontaneous decay per unit time for the level μ_2 is denoted by A, and W is the probability per unit time for the excitation by pump radiation. For a Lorentzian line shape and for transitions induced by a perfectly monochromatic signal, the cross section for stimulated emission is given by

$$\sigma = \frac{c^2}{4\pi^2 \nu^2 \Delta\nu t_s} \tag{3-3}$$

where $\Delta\nu$ is the full width at half maximum of the fluorescence line. A schematic diagram showing the relevant transitions is shown in Figure 3-3. These equations are a starting point for understanding some of the more important properties of solid and gaseous lasers.

Modulation with External Electric Field

The emission of a ruby laser has been modulated by applying an external electric field parallel to the c-axis of the ruby (Reference 3-31). The effect of the electric field is to change the line shape of the fluorescence

and thus change the laser dynamics. The threshold for laser oscillation is obtained by requiring that $d\rho/dt \geq 0$ in Equation 3-1. This is satisfied if

$$\mu_2 - \mu_1 \geq \frac{1}{c\sigma t_2}$$

where we have neglected the degeneracy factors g_1 and g_2.

With Equation 3-3, the relation

$$\mu_2 - \mu_1 \geq \frac{4\pi^2 \nu^2 \Delta\nu}{c^3} \frac{t_s}{t_L} \tag{3-4}$$

which gives the oscillation threshold, is obtained. By changing the line shape ($\Delta\nu$ for example), the threshold for oscillation is changed and more generally the dynamics of the laser will change if σ is caused to vary by the application of an electric field or any other means as seen in Equation 3-4. The results of applying a 40 kv/cm electric field at 10^5 cps to a pulsed ruby laser are quite striking. Figure 3-4 shows an oscillogram of the emission as given in Reference 3-31.

The ruby output is changed drastically by the application of the ac electric field. The normal irregular oscillations of the ruby now appear as approximately constant amplitude spikes appearing periodically with twice the 10^5 cps frequency as the applied electric field and with increased peak power in each spike. The rise of each periodic spike is very sharp (considerably less than a microsecond). The periodicity, the fast rise of the spike and the increased intensity of the spiking, can be explained with the aid of Equations 3-1 and 3-2.

Response Time of Ruby Laser

For the ruby, a three-level laser system, these equations can be written

$$\frac{d\rho}{dt} = \alpha\mu\rho - \frac{\rho}{t_L} \tag{3-5}$$

$$\frac{d\mu}{dt} = -2\alpha\mu\rho - (W + A)\mu + (W - A) N \tag{3-6}$$

where $\mu = \mu_2 - \mu_1$ is the population difference between excited and ground states and $N = \mu_1 + \mu_2$ is the number density of chromium ions, and $\alpha \overset{\text{def}}{=} c\sigma$. The values of ρ and μ, for which time independent operation can

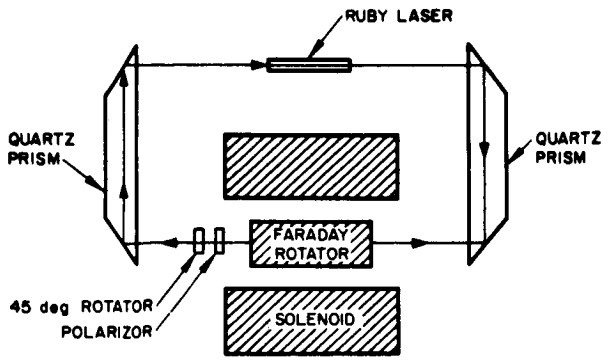


Figure 3-1. Solid-State Laser with Optical Isolator

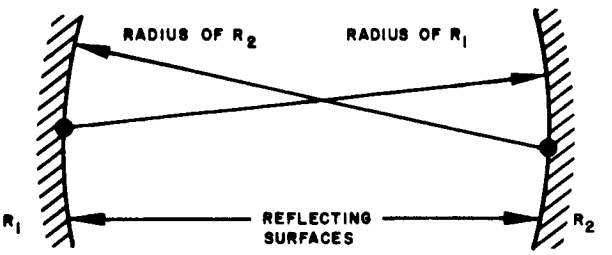


Figure 3-2. Confocal Reflectors Configuration

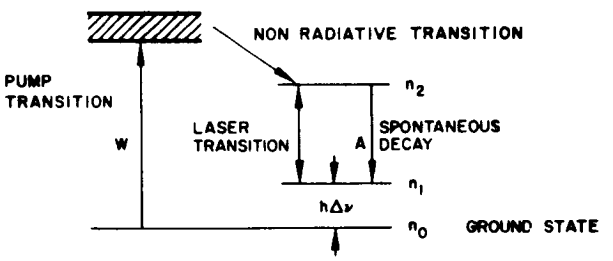


Figure 3-3. Schematic of Laser Transitions

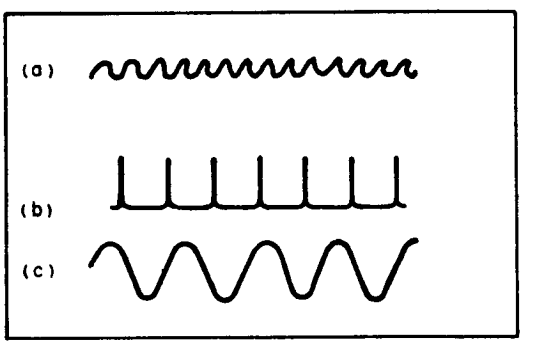


Figure 3-4. (a) Ruby Laser Emission without Electric Field, (b) Emission with 100-kc Field Shown in (c)

occur, are the singular points of this system of equations and are determined by setting $d\rho/dt$ and $d\mu/dt$ simultaneously equal to zero and are thus given by the solutions of

$$\alpha\bar{\mu}\bar{\rho} - \frac{\bar{\rho}}{t_L} = 0$$

$$-\alpha\bar{\mu}\bar{\rho} - (W + A)\bar{\mu} + (W - A)N = 0 \quad , \tag{3-7}$$

where the singular points are denoted by $\bar{\rho}$ and $\bar{\mu}$.

Thus

$$\bar{\mu} = \frac{1}{\alpha t_L} \tag{3-8}$$

$$\bar{\rho} = \frac{A}{2\alpha}(1 + \alpha N t_c)\left(\frac{W}{W_c} - 1\right) \tag{3-9}$$

where

$$W_c = A\left(\frac{1 + \frac{1}{\alpha N t_L}}{1 - \frac{1}{\alpha N t_L}}\right) \tag{3-10}$$

is the threshold pump transition probability per unit time for laser oscillation. Equation 3-8 gives the value of the population density differences for which the cavity has zero gain; this is the threshold population inversion for laser oscillation. Expanding the right-hand side of Equations 3-5 and 3-6 in a power series about the singular point $(\bar{\rho}, \bar{\mu})$ and neglecting higher order terms in $\delta\rho = \rho - \bar{\rho}$ and $\delta\mu = \mu - \bar{\mu}$, the set of linearized dynamical equations is obtained

$$\frac{d\delta\rho}{dt} = \alpha\bar{\rho}\delta\mu$$

$$\frac{d\delta\mu}{dt} = -2\alpha\bar{\mu}\delta\rho - (2\alpha\bar{\rho} + W + A)\delta\mu \tag{3-11}$$

The characteristic equation of the system (Equation 3-11) is

$$\det\begin{pmatrix} -\lambda & \alpha\bar{\rho} \\ -2\alpha\bar{\mu} & -(2\alpha\bar{\rho} + W + A) - \lambda \end{pmatrix} = 0$$

or

$$\lambda^2 + (2\alpha\bar{\rho} + W + A)\lambda + 2\alpha^2\bar{\mu}\bar{\rho} = 0 \qquad (3\text{-}12)$$

The general solution of Equation 3-11 is

$$\delta\rho = a_1 e^{\lambda_1 t} + a_2 e^{\lambda_2 t}$$

where λ_1, λ_2 are roots of Equation 3-12 and a_1 and a_2 are arbitrary constants. For typical values of the parameters α, t_L, W, A, and N the roots of Equation 3-12 will be complex with negative real part (the real part is always negative since α, $\bar{\rho}$, W, and A are never negative). This gives a damped oscillation around the singular point ($\bar{\rho}$, $\bar{\mu}$) which is then a stable focus of the oscillation. When the damping is small the oscillation frequency is

$$\omega \cong \sqrt{2\alpha^2\bar{\mu}\bar{\rho}}$$

or with Equations 3-8 and 3-9

$$\omega \cong \sqrt{\alpha NA\left(\frac{W}{W_c} - 1\right)} \qquad (3\text{-}13)$$

where we have neglected unity compared with $\alpha N t_L$, which is a valid approximation in ruby. Thus ω as given by Equation 3-13 is the frequency of "small" oscillations of the output of the ruby laser. The frequency of spiking observed in many experiments and the frequency given by Equation 3-13 are in reasonable agreement, but the expected damping as predicted by the solutions of the linearized equations is not observed in normal pulsed laser operation or in the single case of CW operation; however, it has been observed in ruby for single mode oscillation.

Under usual conditions $\omega/2\pi$ is of the order of megacycle and Equation 3-13 shows that since α and A are constants for a given material and N cannot be increased beyond certain limits, that ω is essentially limited by the pump intensity obtainable in the ruby. It would be difficult, even under the best circumstances, to obtain $\omega/2\pi$ exceeding 10 megacycles. The laser system is unable to respond to changes of the electric field, or any internal perturbation greater than the natural frequency of the system; these observations show that the system is limited to internal modulation frequencies in ruby of the order of a megacycle.

Figure 3-4 shows that in the modulated signal there is a noticeable sharpening of the leading edge of the spikes and an increase in intensity of the spikes. These effects are caused by establishing a population in excess of what would be the normal inversion in the excited state. Generally, the

effects of time variation of the electric field result in an unusually large gain in the ruby, giving a large $d\rho/dt$ or a fast rise of the emission, and an increase in peak intensity. In fact, the modulated signal that results from the periodic variation of the electric field is equivalent to a periodic Q switching of the cavity, with the difference that here a variation of a ruby parameter (the cross section for stimulated emission) rather than a cavity parameter occurs.

Pulse Position Modulation

The sharpness of the leading edge of the spikes theoretically makes the ruby emission suitable for pulse position modulation. Because of the sharp leading edge a sufficiently fast detector would be able to distinguish between the arrival of spikes that differ by small time increments. The pulses are triggered by an electric field $E \leq E_c$; there will then be a delay Δt for the radiation density to build up in the cavity followed by the fast rise to peak intensity as shown in Figure 3-5. By a suitable time variation of the electric field (in general, not periodic) the pulses could be triggered at specified irregular intervals. In order to obtain sharp, high-intensity spikes the electric field must change a significant amount during the build-up time of the radiation density in the cavity, which is the familiar switching requirement for Q switched lasers (giant pulse). As previously noted, the spikes cannot be spaced less than $2\pi/\omega$ second apart where ω is the natural oscillation frequency of the laser given by Equation 3-13. The fact that one is limited in the rate at which spikes can be produced can be seen qualitatively without referring to the lineratized oscillation equations. When the radiation density becomes very large, the population inversion is driven below the threshold for oscillation, and the laser spike decays. A time interval that depends on the pumping intensity is needed to reestablish the critical inversion. When the critical inversion is reached, a time t is required for the radiation density to build up and the process is then repeated. These effects combine to limit the spiking period to a time of the order of $2\pi/\omega$.

A pulse position modulation scheme using a GaAs laser as the source has been considered in a previous report (Reference 3-32), where some of the merits of pulse position modulation were discussed. A few of the simpler features of essentially the same modulation method with the ruby laser as the coherent radiation source are considered. The minimum sampling period of a signal waveform is set by the natural frequency of the spiking of the ruby laser which is the order of a microsecond. A rise time of the spikes of 10^{-8} second is assumed so about 100 distinguishable positions could be obtained in a sampling interval.* Under these conditions the relative position in a sampling interval can be made a function of the amplitude of the sampled waveform (see Figure 3-6).

*In Reference 3-32 the sampling period and the number of resolution elements (distinguishable positions) was set by the frequency with which the pulses could be repeated and the duration or rise of a single pulse. The estimates were 10^7 pps of $\sim 10^{-9}$ second duration. The number of resolution elements was then taken to be 100.

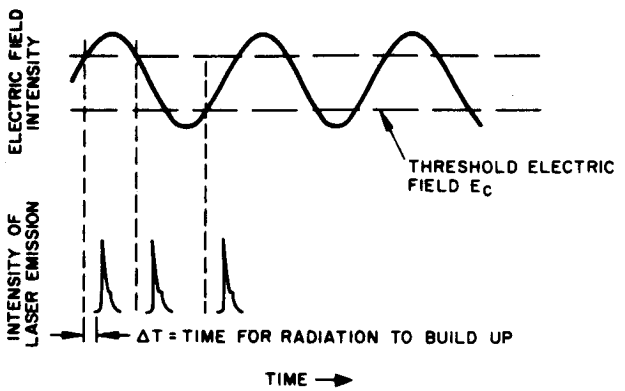


Figure 3-5. Pulse Control by Variable Electric Field

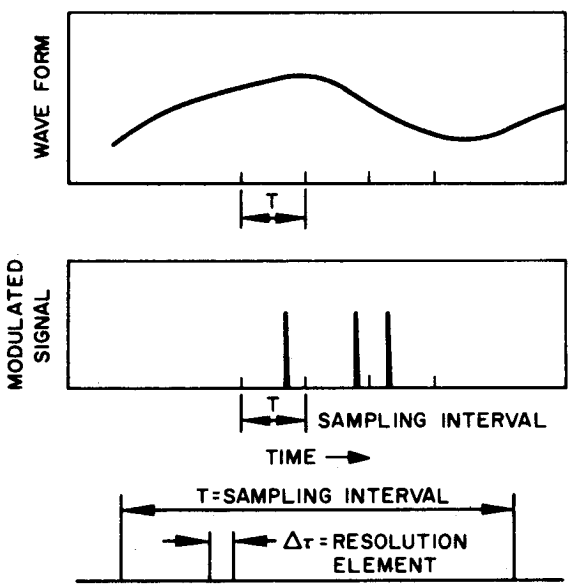


Figure 3-6. Position of Laser Spike as Function of Amplitude of Sampled Waveform

The information per sampling interval is $I = \chi n_2 T/\Delta T$, and the information rate is $1/T \chi n_2 T/\Delta T$. Because of the difference in sampling interval assumed here and for GaAs in Reference 3-9, the information rate obtained for the pulsed GaAs laser is better by a factor of 10 than the ruby laser. A comparison of the lasers is not being made here since the merits of the two should be considered in the context of the system in which they would be used. Furthermore, the performance assumed for both GaAs and ruby has not been realized and it is not known what their future performances will be, especially in regard to continuous or average power.

Giant Pulsing

A particular internal modulation method has been discussed where an electric field parallel to the c-axis of a ruby causes a change in the line shape of the fluorescence. This change has the effect of changing the internal gain of the ruby, making it possible to control the intensity and position in time of the spikes. It was noted that the very fast rise of these modulated spikes might make the controlled spiking output suitable as a carrier signal for pulse position modulation.

The characteristics of the spikes, that is, the control of the time of spiking, the increase in intensity, and the fast rise of the spikes are typical of any giant pulse mode of operation. For example, the first giant pulse operation of ruby was obtained by switching a Kerr cell which effectively varied the reflectivity of the cavity in a time that was small compared with the build-up time of the radiation in the cavity (Reference 3-33). A single spike was obtained that had all the characteristics of the spikes shown in Figure 3-4 for the ruby perturbed by an external electric field. In principle, successive giant pulses similar to those shown in Figure 3-4 may be obtained by repeated switching of a Kerr cell or another Q switching method.

There is an additional method of varying the internal gain of a ruby that results in repetitive controlled pulses. A laser oscillation similar to the one shown in Figure 3-4 has been obtained by applying a time varying magnetic field to the ruby (Reference 3-34). Zeeman splitting of the R_1 line by an inhomogeneous field causes a broadening of the fluorescence line with a resulting control over the gain of the ruby. The experiments were performed at 77°K and the input energy to the flash lamp was about 1.4 times threshold.

The magnetic field required to modulate the signal effectively goes up significantly with temperature. The limitations on the minimum spiking period are essentially the same for Q switching as for internal modulation with electric and magnetic fields.

Giant pulses have also been obtained by Q switching methods in a neodymium glass laser (Reference 3-35). The experiments are currently in progress and the details have not been published. Preliminary results reported are 1.5 joules in a single spike with peak power of 30 megawatts. Such high power will probably not be suitable for the communications applications considered here, but it is possible to regulate the power per spike by

the particular Q switching and pumping intensity employed. Neodymium glass lasers have many desirable properties such as high efficiency (3.9 percent) and energy output of more than 100 joules in a pulse of about 1 millisecond.

Four-Level Lasers

Four-level laser operation occurs when the energy of separation of the terminal laser level and ground state is much larger than the thermal energy of the atom, i.e., when $h\Delta\nu >> kT$ (see Figure 3-3). The four-level crystal materials that have been operated continuously are Nb^{3+}: $CaWO_4$, Dy^{2+}: CaF_2, and U^{3+}: CaF_2.

Continuous laser action in U^{3+} : CaF_2 at 2.613 microns has been obtained, with an emission of about 10 microwatts (Reference 3-36). Threshold for oscillation is slightly less than 2000 watts when cooled by a continuous flow of precooled liquid oxygen. A confocal resonator geometry was used with the silvered face of the crystal 1 percent transmissive. The output is a damped oscillation, the damping occurring in a few milliseconds. There are strong low-frequency fluctuations in the emission that are caused by an ac ripple in the dc power supply. For power levels barely exceeding threshold, a small change in the pumping intensity causes a large fluctuation in the laser output.

Continuous laser action in Nd^{3+} : $CaWO_4$ has been obtained by Johnson et al (References 3-37 and 3-38). The following characteristics are noted: emission occurs at the 1.065-micron line which has the lowest threshold for laser action; the crystal is cooled by a continuous flow of precooled liquid oxygen; the rod is long compared with its diameter and has approximately confocal spherical ends; threshold flash tube power for continuous stimulated emission is about 1300 watts; there is a periodic fluctuation at 60 cps of the power output caused by an ac ripple on the dc power supply. The high frequency spikes of the type that persist indefinitely in the CW operation of ruby are damped out; power output at 1600 watts is about 1 milliwatt superposed on a dc output of 1 milliwatt, and continuous emission is obtained for about 12 seconds. The ac ripple component of the power supply is about 20 percent of the total power to the flash tube (see Figure 3-4 and Reference 3-37)

The signal shown in that article is an example of pump power modulation, a method that has been used successfully for kilocycle modulation of the emission of a Ne - Ne gas laser, and transmission of audio over short distances. The frequency with which a laser signal can be modulated by changing the pump power is limited by the natural frequency of the system as it is when any other parameter that changes the dynamics of the laser oscillation is perturbed. The response of the flash tube, if it is slower than the oscillation frequency of the laser, would then limit the modulation frequency.

Spiking Frequency in Four-Level Systems

The expression for the spiking frequency of a four-level laser is obtained as follows. Because of the difference of the energy of the terminal

laser level and the ground state the population density of the ground state will be negligibly small compared with the population μ_2 of the excited level during the laser oscillation, and the number of atoms in the excited state will be small compared with the total ion number density. Thus Equations 3-1 and 3-2 can be approximated by

$$\frac{d\rho}{dt} = \alpha\mu_2 p - \frac{\rho}{t_L} \tag{3-14}$$

$$\frac{d\mu_2}{dt} = -2\alpha\mu_2\rho + WN - A\mu_2 \tag{3-15}$$

The singular points are

$$\bar{\rho} = \frac{A}{\alpha}\left(\frac{W}{W_c} - 1\right) \tag{3-16}$$

$$\bar{\mu} = \frac{1}{\alpha t_L}$$

where

$$W_c' = \frac{A}{\alpha N t_L} \tag{3-17}$$

W_c' is the threshold pump transition probability per unit time for oscillation of a four-level laser. Linearizing Equations 3-14 and 3-15 as before (see Equation 3-10), the following is obtained

$$\frac{d\delta\rho}{dt} = \alpha\bar{\rho}\,\delta\mu - \frac{\delta\bar{p}}{\tau_L}$$

$$\frac{d\delta\mu_2}{dt} = -\alpha\bar{\mu}_2\,\delta\rho - (\alpha\bar{\rho} + A)\,\delta\mu_2 \tag{3-18}$$

The characteristic equation is

$$\lambda^2 + (\alpha\bar{\rho} + A)\,\lambda + \alpha\bar{\mu}_2\bar{\rho} = 0$$

The angular frequency of the damped oscillation when the damping is small compared with the oscillation frequency is approximated by

$$\omega \cong \sqrt{\frac{A}{t_L}\left(\frac{W}{W'_o} - 1\right)} \qquad (3\text{-}19)$$

where we have used Equations 3-16 and 3-17. The spontaneous emission time for Nd^{3+}: $CaWO_4$ is $\sim 10^{-4}$ second, so $A = 10^4$ sec^{-1}. Take $t_L = 5 \times 10^{-9}$ for the decay time of the radiation in the cavity, and let $W/W'_o = 2$. Then Equation 3-19 gives $\omega \approx 1.4 \times 10^6$ sec^{-1} and $\omega/2\pi = 2.2 \times 10^5$ cps which is in fair agreement with the observed oscillation frequency of Nd^{3+}: $CaWO_4$ operated under conditions similar to those assumed above. (For example, see Reference 3-16 and Figure 3-3.)

CW Operation in Nd^{3+}: $CaWO_4$ and Dy^{2+} CaF_2

More recently Johnson (Reference 3-38) has obtained about 0.5 watt continuous power from Nd^{3+}: $CaWO_4$ at 1500 watts in a confocal resonator geometry. Measurements of spectral width, directionality, power output, and spiking behavior were made and near-field photographs of the intensity distribution over the face of the laser rod were taken. Experiments were made in both conical and Fabry-Perot geometry. A connection between the number of resonator modes excited and the spiking behavior has been observed.

Continuous laser action takes place at 85°K when the crystal is cooled by a continuous flow of liquid oxygen that has been precooled to liquid nitrogen temperature in order to eliminate bubbling. The emission consists of the single line with a spectral width of 0.2 micron which is a reduction of the natural fluorescence line width by a factor of 30. Beam width in the Fabry-Perot geometry is about 0.2 to 0.3 degree, while divergence of the emission of the continuously operating laser with confocal geometry is much larger.

Near field patterns of the light emitted from the end faces of the crystals with Fabry-Perot geometry indicate that a small number of resonator modes are excited, and that the number of active regions of the crystal is essentially independent of the pumping power. However, in the confocal geometry near-field patterns show a high degree of symmetry with the number of excited modes increasing as the power is increased. At a power level of 1.3 times threshold a large number of resonance modes appear to be excited. Spiking or undamped oscillations persist in continuous operation in the Fabry-Perot geometry for all values of input power, while in the confocal geometry the continued spiking disappears as the power level increases, and at power levels well above threshold essentially no spikes are observed. Since the near-field patterns indicate many resonator modes are excited at high power input, Johnson has suggested there may be a connection between mode structure and continued spiking in continuous operation (Reference 3-39).

Continuous operation in Dy^{2+}: CaF_2 has been obtained by Kiss and Duncan (Reference 3-40), L. F. Johnson (Reference 3-41), and A. Yariv (Reference 3-42). The strongest emission line is at 2.360 microns; laser

action in CW operation occurs at this line. In contrast to the CW operation of Nb^{3+}: $CaWO_4$, where a dynamic cooling technique was needed, the dysprosium laser is cooled to 20° K by immersion in liquid hydrogen. The laser was excited by a flash lamp driven by a 120-cps power supply and also by a dc-power supply. Threshold for CW operation was 650 watts rms input power to the lamp. Another crystal tested had a threshold at 350 watts at 20° K. The peak output power at three times threshold was 2.5 milliwatts.

Yariv obtained CW emission in Dy^{2+}: CaF_2 using both dynamical and immersion cooling. Measurements of threshold power were made for a number of different concentrations of Dy at both ac and dc excitations. The results obtained for continuous operation at 77° K are given below.

Doping, percent Dy^{2+} CaF_2	Ac Threshold, watts	Dc Threshold, watts
0.5	800	800
0.2	600	660
0.1	300	350
0.05	180	210
0.02	80	100

The 80- and 100-watt thresholds are the lowest yet obtained in CW operation of solid materials. The maximum power output was obtained with 0.05 percent concentration of Dy with a highly transmissive slivered end mirror; 0.3 watt was obtained with 800 watts input to the lamp. Continuous runs up to 30 minutes were maintained with no adverse effects. Spiking was observed to decay in about 0.3 m /sec after the onset of oscillation.

It should be noted that the 2.36 micron line is not absorbed by water vapor which is the main source of absorption in the atmosphere.

In communications applications where a high power constant amplitude carrier signal is desired (because of the damped spiking characteristics and relatively high power) Nb^{3+}: $CaWO_4$ and Dy^{2+}: CaF_2 appear to be promising laser materials. However, it should be remembered that the oscillations are damped at the expense of a larger beam angle.

SEMICONDUCTOR LASERS

The recent achievement of laser action in semiconductor diodes has given rise to a new field of investigation in the realm of laser technology. Its significance stems from the fact that direct conversion of electrical energy to coherent radiation takes place in a highly efficient manner.

Additionally the laser diode provides a conceptually simple method of modulating a laser beam since the electrical pumping current can easily carry the modulating signal as well.

Operating Characteristics

Current passing through a GaAs diode in the forward direction gives rise to the emission of recombination radiation in the vicinity of the p-n junction. As the current density is increased, the width of the emitted spectral line is reduced indicating the presence of stimulated emission. By controlling crystal impurities and creating an optical cavity in the plane by the junction, laser action can occur. It is believed that at high current densities operation takes place with nearly unity quantum efficiency, i.e., for every electron crossing the junction one photon is emitted. The net conversion efficiency of the device (input electrical energy to radiated energy) has been observed to be as great as 50 percent compared to the present 0.1 percent for gaseous or solid-state lasers. Clearly, these highly efficient devices merit further investigation.

Although experimental results are sparse, some idea of the operating characteristics of present semiconductor lasers can be obtained from the partial summary given in Table 3-4.

Coherent light emission has now been observed from several different forward-biased injection diodes. As in solid and gaseous lasers, the onset of laser action in the injection diodes is characterized by a narrowing of the spectrum of the emitted radiation (see Figure 3-7). Laser action has been obtained in GaAs (References 3-43, 3-44, 3-45), Indium Arsenide (Reference 3-46), Indium Phosphide (Reference 3-47), and Ga $(As_{1-x}P_x)$ (Reference 3-48) diodes. Laser action is obtained most easily at low temperatures in a pulsed mode of operation, however, continuous operation has been obtained in GaAs and InAs, and pulsed operation of GaAs at room temperature has been achieved.

Power and Efficiency

In the pulsed mode of operation the diode laser is excited directly by current pulses of less than a microsecond duration. The inherently low capacity of these diode systems in principle permits current pulses of the order of a nanosecond with the subsequent emission of nanosecond pulses of coherent radiation. One of the desirable features of semiconductor lasers is that the response time of the laser is fast enough to allow relatively high (10^9 - 10^8 cps) direct modulation frequencies. By way of comparison, the natural frequency of present solid-state lasers is in the range of 10^6 to 10^7 cps.

Current densities of the order of 10^3 amps/cm^2 through the diode are required for laser oscillations. The area of the p-n junction is typically about 10^3 cm^2 and the emitting layer is about 15 microns thick (Reference 3-49). The face that the emitting layer is so thin means that a slight amount of dissipation of energy is likely to cause a serious heating problem. This,

coupled with the fact that a rise in temperature causes a degradation of laser performance, is one of the factors that presently limit the power output of semiconductor lasers. The total average light output of about 7 milliwatts from a pulsed GaAs laser was reported (Reference 3-50). The output consisted of 1000 pps of 10 microseconds duration. The efficiency of the laser was measured to be less than 30 percent. Current pulses 0.4 microsecond long with a repetition rate of 200 per second were used with a InAs laser (Reference 3-46). The highest current applied to the diode was 60 amperes. With the duty cycle implied above (8 by 10^{-5}) the total average power out cannot be more than 10 milliwatts. An InAs diode was operated continuously with a magnetic field of 4.1 K gauss, a temperature of 4.2° K and 210 milliamperes dc current in the diode. Another InAs diode was operated continuously at 2.1° K and currents up to 1 ampere (Reference 3-51). One of the effects of the magnetic field was to lower the threshold current for oscillation. The total light emission was not reported. Higher power outputs have been reported in some of the trade journals recently, but details of their operation have not yet been published.

The future power capabilities of diode injection lasers have not been determined. It should be noted that all of the emitted power of the diode lasers may not be useful in a communication channel. For example, in the continuous operation of GaAs (Reference 3-52) only about 10 percent of the total radiation was coherent. Melngailis, however, has reported that almost all of the CW emission of the InAs diodes was coherent. In addition, problems similar to those encountered in the high-power operation of other laser materials occur also with the diode injection lasers. For example, as power input (or current) is increased above threshold for oscillation many modes are excited. Often these modes are strongly off axis and can result in large beam divergences. In general it is not likely that conditions for which high power output is obtained will be conducive to good spectral and directional properties of the beam.

The efficiency of GaAs injection lasers have been studied by G. Cheroff et al. (Reference 3-50). Two efficiencies have been defined, the external quantum efficiency, which is the ratio of the rate with which photons are emitted to the rate of electron flow through the junction, and the internal quantum efficiency, which is defined as the number of photons generated per injected electron. In tests on a series of diodes operated at liquid nitrogen temperature, the internal and external quantum efficiencies have been measured. The diodes were driven by microsecond pulses at the rate of 1000 pps.

The external quantum efficiency, which is essentially the overall power efficiency, increases as the current is increased above threshold. At currents about three times threshold the overall efficiency was about 26 percent. The internal quantum efficiency for directional emission was 70 percent which is considerably lower than the 100 percent reported by Quist et at. (Reference 3-51). The internal quantum efficiency is important since it places an upper limit on diode laser efficiency.

TABLE 3-4. CHARACTERISTICS OF SEMICONDUCTOR LASERS

Material	Manufacturer	Mode	Temperature °K	Threshold, amp/cm^2	Power	Beam and Spectral Characteristics
GaAs	IBM	Pulsed		10,000		Two lines observed 6 Å apart and 2 Å wide 4-degree beam divergence
GaAs	IBM	Continuous	4.2	100	10 to 25 milliwatt output for 50 milliwatt input	8400 Å
GaAs	Philco	Continuous	300	–	1 milliwatt output for 120 milliwatt input	~9000 Å
GaAs		Continuous	77	–	25 milliwatt output for 120 milliwatt input	~9000 Å
GaAs	GE	Pulsed	77	8,500	–	Coherence over 100 microns 0.1-degree beam divergence
GaAs		Pulsed	4.2	–	–	Line narrows from 125 Å to less than 15 Å
GaAs P 1-x x		Pulsed	77	Less than 11,000	1 to 5 microseconds input pulses 1.3 volts at 20 to 100 milliamperes	Line narrows: 125 Å→12 Å 7100 Å

TABLE 3-4 (continued)

Material	Manufacturer	Mode	Temperature °K	Threshold, amp/cm^2	Power	Beam and Spectral Characteristics
GaAs	MIT	Pulsed	77	10,000	-	Less 10-degree beam divergence
					5 micro-second pulses at 13 cps input	Line narrows: 175 Å→30 Å 100 Å→5 Å
GaAs		Pulsed	4.2	700	280-watt peak output	Multiple modes observed

The inherently high efficiency of injection diode lasers and the high frequency direct current modulation methods are attributes not shared by other lasers.

Directional Properties

G. Burns et al (Reference 3-53) have fabricated GaAs junctions that show strong directionality effects. The angular distribution of the emitted intensity has been related to some simple geometric standing wave modes. The best performance was obtained in a parallel-piped structure of dimensions $\sim 5 \times 10^{-3}$ by 50×10^{-3} inches, with the long dimension of the diode in the plane of the p-n junction. Laser oscillation at 8400 degrees first occurs in the direction of the long dimension of the cavity with about 8 amperes through the junction. Laser oscillations in the short direction are excited at about 18 amperes. Laser action was obtained at liquid nitrogen temperature with the diode driven by current pulses of 100 nanoseconds duration. It was not necessary to coat the reflecting surfaces perpendicular to the long direction with any reflective substance because the high index of refraction of GaAs provided sufficient reflectivity (35 percent) to obtain laser oscillation.

For currents slightlyabove 8 amperes the beam angle of the coherent radiation in a plane parallel to the junction plane is approximately 2.5 degrees. The width of the junction plane is 1.3 by 10^{-2} cm and if the diffraction limit is calculated for this dimension and the 8400 A wavelength, the beam angle is within about a factor of 10 of the diffraction limit. As the current through the diode is increased above threshold, additional peaks in the angular distribution appear as shown in Figure 3-8 and the beam angle consequently increases. Similar intensity distributions are given in Reference 3-53 for several values of diode current.

The positions of the peaks in the intensity distribution have been correlated with some simple standing wave patterns. For example, a peak in the intensity has been observed at about 25 degrees and this peak would correspond to the standing wave shown in Figure 3-9. The angular distribution of the emitted intensity in the plane perpendicular to the junction plane was qualitatively similar in width to the angular distribution in the plane of the junction.

Below threshold the spectral width at half maximum of the incoherent emission, at liquid nitrogen temperature, is 125 A. As the current is increased through threshold, a sudden decrease of line width to about 15 A has been observed (see Figure 3-7). This 15-A emission is composed of several quite narrow unresolved lines. Recently the spectral width of the distinct lines has been observed to be 0.2 A and one observation showed a line width of 0.05 A, which was the resolution limit of the instrument (Reference 3-54).

The laser emission of InAs at 3.1 microns shows spectral characteristics similar to GaAs. Above threshold a relatively narrow line appears out of a broad spontaneous emission line of half width 1900 A at 77° K. The spectrum showed lines with a half width of 35 A, with the number of lines

increasing at higher currents. The beam angle in the plane parallel to the junction was about 5 degrees which is comparable to that observed in GaAs.

Indium phosphide lasers emit between 9030 and 9100 Å for most of the diodes tested (Reference 3-47). The full line width of the spontaneous emission at 77° K is between 140 and 180 Å and between 65 and 100 Å at 4.2° K. At 77° K the spectrum of the emission consists of several lines of full width at half maximum of about 1 Å and the lines are separated by about 3.5 Å. When operated at 4.2° K, lines of 0.5 Å were observed which was the resolution limit of the measurement.

In diode lasers the wavelength of the spontaneous emission that gives rise to the laser oscillation is an increasing function of temperature. This causes the wavelength of the laser emission and the spacing of the modes to vary with temperature. The assumption that the cavity resonances are standing wave modes in Fabry-Perot geometry gives results that are consistent with observations on mode spacing and the temperature dependence of the laser emission lines in GaAs and InP diodes.

The wavelength of the Fabry-Perot modes are given by

$$\lambda = \frac{2\mu\ell}{m} \tag{3-20}$$

where μ is the index of refraction and is dependent on wavelength and temperature, ℓ is the distance of separation of the reflecting end faces of the cavity and depends on temperature, and m is an integer.

To obtain an expression for $d\lambda/dt$ we differentiate Equation 3-20 with the result

$$\frac{d\lambda}{dt} = \left(\frac{\partial\lambda}{\partial\mu}\right)_{\ell} \left[\left(\frac{\partial\mu}{\partial T}\right)_{\lambda} + \left(\frac{\partial\mu}{\partial\lambda}\right)_{T} \frac{d\lambda}{dT}\right] + \left(\frac{\partial\lambda}{\partial\ell}\right)_{\mu} \frac{d\ell}{dT}$$

Using Equation 3-20 to obtain $\left(\frac{\partial\lambda}{\partial\mu}\right)_{\ell}$ and $\left(\frac{\partial\lambda}{\partial\ell}\right)_{\mu}$ and substituting it in the above expression gives

$$\frac{d\lambda}{dT} = \frac{\ell\left(\frac{\partial\mu}{\partial T}\right)_{\lambda} + \mu\frac{d\ell}{dT}}{1 - \ell\left(\frac{\partial\mu}{\partial\lambda}\right)_{T}} \tag{3-21}$$

G. Burns et al (Reference 3-55), using the rough experimental values of $(\partial\mu/\partial T)_{\lambda}$, $(\partial\rho/\partial\lambda)_{T}$, and $d\ell/dT$ obtained by T.M. Quist, et al (Reference 3-51), find $d\lambda/dT$ from Equation 3-21 to be 0.4 Å/deg at room and liquid

nitrogen temperatures, which is in agreement with the observed rate of change of wavelength of the individual lines with temperature at 77° and 300° K.

The wavelength of the spontaneous emission changes at the rate 2.5 Å/deg at 77° K. Since $d\lambda/dT$ for the stimulated emission is 0.4 Å/deg the stimulated emission lines do not follow the shift in the spontaneous emission. Thus as the temperature is changed, lines will disappear and new lines will appear. The shift in wavelength with temperature is probably one of the causes of line broadening in pulsed operation.

An approximate expression for mode spacing can be obtained by taking the differential of (1) with ℓ and T constant. Thus,

$$\Delta\lambda \approx \left(\frac{\partial\lambda}{\partial\mu}\right)_m \frac{d\mu}{d\lambda}\ \Delta\lambda + \left(\frac{\partial\lambda}{\partial m}\right)_\ell \Delta m$$

and with $\Delta m = -1$

$$\Delta\lambda \approx \frac{\lambda^2}{2\mu\ell} \frac{1}{\left(1 - \frac{\lambda}{\mu}\frac{d\mu}{d\lambda}\right)} \qquad (3\text{-}22)$$

The separation, $\Delta\lambda$, of adjacent modes depends on temperature through the temperature dependence of λ, ℓ, μ, and $d\mu/d\lambda$. Measurements of mode separation made by Burns and Nathan with a GaAs laser operating at room temperature are in agreement with values of $\Delta\lambda$ calculated from Equation 3-22. I. Melngailis and R. H. Rediker have also observed mode separation (in the continuous operation of InAs in a magnetic field) that is in essential agreement with Equation 3-22.

Threshold for oscillation is a strong function of temperature (Reference 3-56). Measurements in a GaAs diode show about a third power dependence on temperature above 60° K, and below 60° K the curve flattens out at low values of threshold current density as shown in Figure 3-10. The small threshold currents at low temperatures are the reason why continuous operation is obtained more easily at low temperatures.

The first CW GaAs laser was operated at 2.0° K and oscillation could not be obtained at 2.1° K. Recently continuous laser emission has been obtained at 77° K with the two optically flat surfaces coated with silver.

The oscillation threshold will be more sensitive to a change in reflectivity when the dimensions of the cavity are small. As in the solid and gaseous lasers, the threshold condition depends on losses due to scattering and absorption in the material and transmission through the reflecting surfaces. Lasher (Reference 3-57) has given an expression for threshold current density, j, where the threshold current is proportional to the cavity losses. Thus

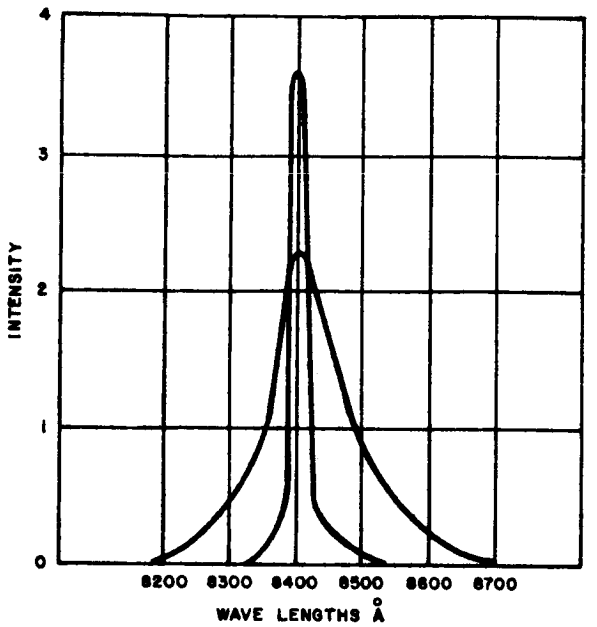


Figure 3-7. Spectral Distribution from GaAs p-n Junction Diode Below and Above Threshold (Different vertical scales)

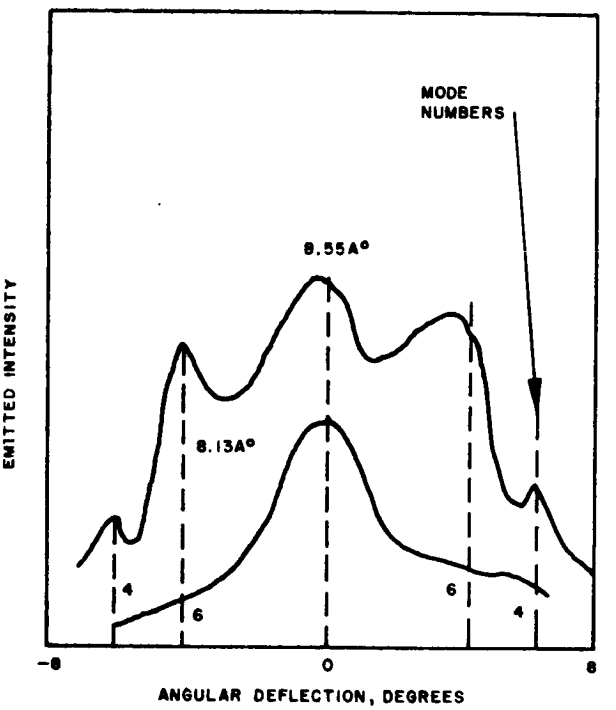


Figure 3-8. Angular Distribution of Emitted Intensity in Plane of p-n Junction

Threshold = 8.13 amperes

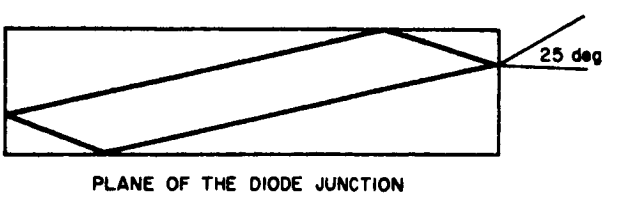


Figure 3-9. Standing Wave Mode Emitting Off-Axis Radiation

$$j \propto \left(\frac{1}{t_r} + \beta\right)$$

where $1/t_r \approx (1 - R)\, c/\ell$ accounts for the transmission losses, where R is the reflectivity and ℓ is the length of the cavity and c is the velocity of light. The other losses are described by the single parameter β. We see that if the transmission loss term is dominant, which may happen for very small ℓ, then the threshold current will be sensitive to changes in R. If the absorption and scattering losses are large compared with the transmission losses then the threshold current will not be reduced significantly by increasing the reflectivity R. For units having a typical length $\approx 10^{-2}$ cm the threshold has been reduced as much as a factor of four by increasing the reflectivity.

GASEOUS LASERS

Due to the excellent survey article on gas lasers by Bennett (Reference 3-58) and since the statistical properties of the He - Ne laser are discussed in Section 4, only a brief summary of gaseous laser performance is given here. The section is then concluded with an analysis of the limitation on gas laser power for direct atom-atom collision excitation.

Gaseous Laser Characteristics

Laser oscillations of frequencies ranging from the visible (0.63 micron) to the middle infrared (12 microns) have been obtained in gaseous materials. So far, maser action has been obtained in ten different gaseous systems (see Table 3-5). A considerable amount of theoretical and experimental work has gone into the study of the He - Ne gas system (Reference 3-59). The oscillator consists of a Fabry-Perot interferometer about 1 meter long, and population inversion is achieved by exciting helium atoms to a metastable state in an electrical discharge. The helium atoms transfer the excitation energy to the neon gas by collisions. Under suitable conditions a stimulated transition of the excited neon atoms to a lower level takes place to give the maser oscillation.

The output of gaseous lasers is characterized by extremely monochromatic, narrow-beam, continuous radiation. Inherent line widths of 2 cps have been obtained in a He - Ne system (Reference 3-60). Frequency stability has been found to be 2 parts in 10^9 over 100-second time intervals. It is expected that it will be possible to obtain a high degree of long-term frequency stability in gaseous lasers.

The power output of most gas lasers has been of the order of milliwatts. In the He - Ne system 15 milliwatts was obtained in a single line with an estimated power dissipation of 50 watts. Saturation occurs at about 75 watts and additional power input does not increase the output power in the beam. Bennett (Reference 3-58) suggests that through the use of mode suppression

TABLE 3-5. SUMMARY OF GASEOUS MASER TRANSITIONS AND BEAM POWER

Gas	Wavelength, microns	Continuous Power, mw per beam	Reference (first work)
Helium-neon	0.6328	4	3-26
	1.0798		
	1.0845		
	1.1143		
	1.1177		
	1.1390		
	1.1409		
	1.1523	20	3-16
	1.1601		
	1.1614	1	3-16
	1.1767		
	1.1985		
	1.2066		
	1.5231	3	3-27
	3.3913	10	3-28
Cesium	7.1821	0.025	3-29
Neon-oxygen	0.84462	1	3-30
Argon-oxygen	0.84462	1	3-30
Helium	2.0603	3	3-31
Neon	1.1523	1	3-32
	2.1019	1	3-31
	5.40		3-33
Argon	1.6180		
	1.6941	0.5	3-31
	1.793		
	2.0616	1	3-31
Krypton	1.6900		
	1.6936		
	1.7843		
	1.8185		
	1.9211		
	2.1165	1	3-31
	2.1902	1	3-31
	2.5234		
Xenon	2.0261	5	3-31
	5.5738		3-33

TABLE 3-5 (continued)

Gas	Wavelength, microns	Continuous Power, mw per beam	Reference (first work)
Helium-xenon	2.0261 2.3193 2.6269 2.6511 3.1069 3.3667 3.5070 3.6788 3.6849 3.8940 3.9955 5.5738 7.3147 9.0040 9.7002 12.263 12.913	10	3-31 3-33

techniques in long interferometers, cooling techniques in multiple-tube structures, or power amplifiers, a sizable fraction of a watt might eventually be obtained.

Gaseous Laser Power Considerations

Gas lasers, for reasons that are related to their low density, do not have the same potential for high-power pulsed operation that exists in solid-state devices. An analysis of this limitation, for RF and d c pumping where the excitation mechanism is by means of direct atom-atom collision, is given here. Other methods of excitation, such as direct electron excitation, which may have the potential for higher power, are currently being investigated.

Figure 3-1 shows the energy level diagram for the helium-neon gas laser. The dynamical equations governing the population densities of the neon states are as follows:

$$\frac{d\rho}{dt} = \alpha\rho\ (\mu_2 - \mu_1) - \frac{\rho}{t_c} \tag{3-23}$$

$$\frac{d\mu_2}{dt} = \alpha\rho\ (\mu_2 - \mu_1) + \mu_{He1} < \sigma_a \nu_a > \mu_g - \frac{\mu_2}{t_{s2}} \tag{3-24}$$

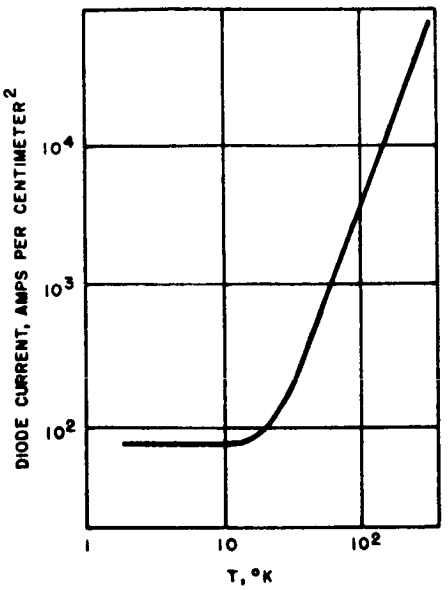


Figure 3-10. Threshold for Stimulated Emission

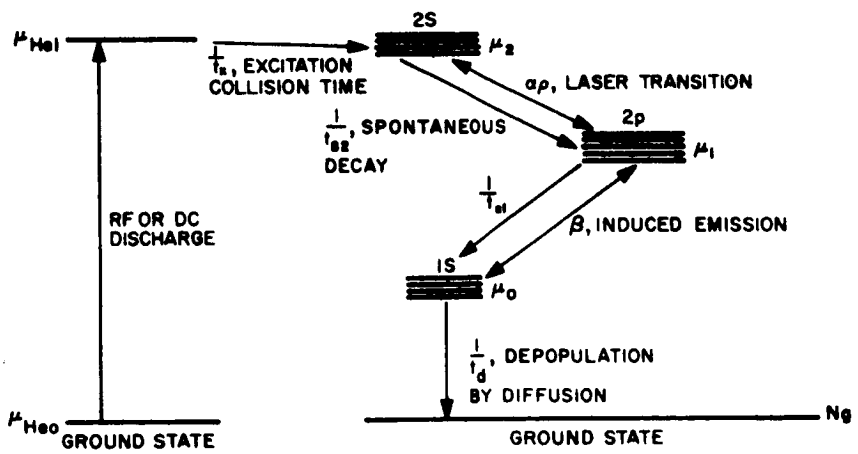


Figure 3-11. He-Ne Gas Laser Energy Level Diagram

$$\frac{d\mu_1}{dt} = a\rho\,(\mu_2 - \mu_1) - \frac{\mu_1}{t_{s1}} - \beta\,(\mu_1 - \mu_o) + \frac{\mu_2}{t_{s2}} \qquad (3\text{-}25)$$

$$\frac{d\mu_o}{dt} = \frac{\mu_1}{t_{s1}} + \beta\,(\mu_1 - \mu_o) - \frac{\mu_o}{t_d} \qquad (3\text{-}26)$$

$$\mu_g + \mu_o + \mu_1 + \mu_2 = N \qquad (3\text{-}27)$$

where

ρ = photon density of coherent radiation

μ_2 = population density of 2S neon state

μ_1 = population density of 2p neon state

μ_o = population density of 1S neon state

μ_g = population density of ground state

μ_{He1} = population density of helium excited state

a = $c\sigma$ (σ = cross section for stimulated 2S - 2p laser transition, c = velocity of light)

t_c = cavity decay time for 2S - 2p radiation

t_d = time for neon atom in excited 1S state to diffuse to wall

t_x = $\mu_{He} \langle \dot{\sigma}_a \nu_a \rangle$ frequency of excitation of neon 2S state per ground state near atom

t_{s2}, t_{s1} = spontaneous decay times of levels 2S and 2p

The transition rate between the neon 2p--1S states is taken to be proportional to the population density difference $(\mu_1 - \mu_o)$. The proportionality factor is denoted by β and depends on the density of 2p--1S radiation in the cavity.

The equations that govern the population of the helium atomic levels have been omitted. These equations are connected to Equations 3-23 to 3-26 through the parameter $\mu_{He1} \langle \sigma_a \nu_a \rangle = 1/t_x$. Thus all the complicated

details of the electric discharge and the population and depopulation of the μ HeI state is collected in the parameter t_x, which is the characteristic time for excitation of the 2S state by atom-atom collisions, are collected.

The term $\beta(\mu_1 - \mu_o)$ in Equations 3-25 and 3-26 accounts for the so-called radiation trapping. Apparently the spontaneous decay time of the IS level is so long that de-excitation at the walls is needed to depopulate the IS state. The population density of the IS level and consequently the population of the 2p and 2S states will therefore be strongly dependent on the diffusion time and hence on the pressure, temperature, and geometry of the discharge tube. The experimental fact that the gain of the helium-neon gas laser is inversely proportional to the diameter of the discharge tube is taken to be evidence of the dependence of laser oscillation on the diffusion process.

Equations 3-23 to 3-26 are approximate for a number of reasons; one of them is that diffusion to the walls has been accounted for only for atoms in the 1S state. This is justifiable if the other relevant processes have smaller characteristic times than the diffusion time. A comparison of these times has not been made, but for the present possible complications that may arise from diffusion effects on the population densities of other levels are ignored.

Time-independent behavior is obtained when the derivatives in Equations 3-23 to 3-26 are simultaneously zero. The values of the dependent variables for which this occurs are denoted by barred symbols. Thus, in steady-state operation,

$$\alpha \bar{p} (\bar{\mu}_2 - \bar{\mu}_1) - \frac{\bar{p}}{t_c} = 0 \tag{3-28}$$

$$- \alpha \bar{p} (\bar{\mu}_2 - \bar{\mu}_1) + \frac{\bar{\mu}_g}{t_x} - \frac{\bar{\mu}_2}{t_{s2}} = 0 \tag{3-29}$$

$$\alpha \bar{p} (\bar{\mu}_2 - \bar{\mu}_1) - \frac{\bar{\mu}_1}{t_{s1}} - \beta (\bar{\mu}_1 - \bar{\mu}_o) + \frac{\bar{\mu}_2}{t_{s2}} = 0 \tag{3-30}$$

$$\frac{\bar{\mu}_1}{t_{s1}} + \beta (\bar{\mu}_1 - \bar{\mu}_o) - \frac{\bar{\mu}_o}{t_d} = 0 \tag{3-31}$$

$$\bar{\mu}_g + \bar{\mu}_o + \bar{\mu}_1 + \bar{\mu}_2 = N \tag{3-32}$$

Let $\mu = \mu_2 - \mu_1$; then solution of the above equations yields

$$\bar{\mu} = \frac{1}{\alpha t_c} \tag{3-33}$$

$$\bar{\mu}_o = \frac{N - \bar{\mu}}{2} \frac{1}{\frac{t_d + t_x}{2t_d} + \left(\frac{1 + \beta t_d}{1 + \beta t_{s1}}\right) \frac{t_{s1}}{t_d}} \tag{3-34}$$

$$\bar{\mu}_1 = \bar{\mu}_o \left(\frac{1 + \beta t_d}{1 + \beta t_{s1}}\right) \frac{t_{s1}}{t_d} \tag{3-35}$$

and

$$\bar{\rho} = \left[\left(\frac{N - \bar{\mu}}{2}\right) \frac{\frac{t_{s2}}{t_{s1}} - \frac{1 + \beta t_d}{1 + \beta t_{s1}}}{\frac{t_d + t_x}{2t_{s1}} + \frac{1 + \beta t_d}{1 + \beta t_{s1}}} - \bar{\mu}\right] \frac{t_c}{t_{s2}} \tag{3-36}$$

which is the equation of interest at the moment.

Equation 3-36 gives the photon density expressed in terms of the characteristic times of the system. Since the inversion required for oscillation is much less than the density N of neon atoms a necessary condition for laser oscillation may be approximated by

$$\frac{t_{s2}}{t_{s1}} > \frac{1 + \beta t_d}{1 + \beta t_{s1}} \tag{3-37}$$

where $t_{s2}/t_{s1} > 1$ in the He - Ne system. From Equation 3-34 it can be seen that if the diffusion time t_d becomes too large the inequality will not be satisfied and oscillation is not possible. Evidently this is the reason laser action cannot be obtained in discharge tubes of large diameter.

When Equation 3-37 is satisfied the inequality

$$\bar{\rho} < \frac{N}{2} \frac{t_c}{t_{s1}} \frac{1}{\frac{t_d + t_x}{2} + \frac{1 + \beta t_d}{1 + \beta t_{s1}}} \tag{3-38}$$

is satisfied and gives an upper bound on the photon density. It is clear from Equation 3-38 that to obtain high-energy density (high power) it is attempted in practice to satisfy strongly the inequalities

$$\frac{t_x + t_d}{t_{s1}} < 1 \quad , \quad \frac{1 + \beta t_d}{1 + \beta t_{s1}} < 1 \quad , \quad \text{and} \quad \frac{t_c}{t_{s1}} > 1 \qquad (3\text{-}39)$$

Equation 3-38 can be simplified slightly in the two limiting cases $\beta t_s << 1$ and $\beta t_s >> 1$. The first case corresponds to the probability per unit time for induced transition between 2p--1S levels being small compared with the spontaneous decay probability. In this case,

$$\bar{p} < \frac{N}{2} \frac{t_c}{t_{s1}} \quad \text{if} \quad \beta t_s << 1 \qquad (3\text{-}40)$$

for the upper bound on the photon density.

If $\beta t_s >> 1$ the induced transition probability per unit time is much larger than the spontaneous decay probability, and the upper bound of the photon density is given by

$$\bar{p} < N \frac{t_c}{t_x + 3t_d} \quad \text{if} \quad \beta t_s >> 1 \qquad (3\text{-}41)$$

The power output is proportional to the photon density and the area of the discharge tube, and is thus proportional to the gas density N. The area is limited by the oscillation condition (Equation 3-37) and this limitation turns out to be quite severe (judging by the areas of discharge tubes used in practice). It can be seen that the photon density is relatively low because of the low N of the gas (compared with solid lasers). The He - Ne gas laser is essentially a low-power device because of the inherently low N and the restriction on size that is expressed by the oscillation inequality (Equation 3-37).

It is of interest to compare these results with the photon density in ruby. The dynamical equations can be written

$$\frac{d\rho}{dt} = \alpha \mu \rho - \frac{\rho}{t_c} \qquad (3\text{-}42)$$

$$\frac{d\mu}{dt} = -2\alpha\mu\rho - (W + A)\mu + (W - A) N \qquad (3\text{-}43)$$

where W is the pump transition probability per unit time, t_c is the cavity loss time, and $A = 1/t_s$ is the spontaneous transition probability per unit time. The time derivatives are simultaneously zero when

$$\mu = \bar{\mu} = \frac{1}{a t_c} \tag{3-44}$$

$$\rho = \bar{\rho} = \frac{N}{2 a \bar{\mu}} \left[\frac{2 A W_c}{W_c + A} \left(\frac{W}{W_c} - 1 \right) \right] \tag{3-45}$$

where W_c is proportional to the minimum threshold pump power and is defined by

$$\frac{W_c - A}{W_c + A} = \frac{\bar{\mu}}{N} \tag{3-46}$$

Since $\bar{\mu}/N << 1$, $W_c \approx A$ and $W_c + A/2 \approx W_c$. Hence $\bar{\rho}$ may be approximated by

$$\bar{\rho} = \frac{N}{2} \frac{t_c}{t_s} \tag{3-47}$$

where Equation 3-44 has been used and $A = 1/t_s$. Comparing Equation 3-47 for ruby with Equations 3-38, 3-40, and 3-41 for the gas laser it is clear why ruby (and solids in general) have inherently higher power capabilities than the He - Ne laser. As was mentioned before, the dominating factor is the difference in the possible doping density N for solids and gases together with the restriction on the size of gas lasers implied by Equation 3-37.

REFERENCES

3-1. H. Kogelnik and C. K. N. Patel, Proc. IRE, 50, 2365 (1962).

3-2. D. R. Herriott, J. Opt. Soc. Am., 52, 31 (1962).

3-3. J. F. Ready, Proc. IRE, 50, 1695 (1962).

3-4. T. H. Maiman, et al., Phys. Rev. 123, 1151 (1961).

3-5. B. J. McMurty and A. E. Siegman, Appl. Opt., 1, 51 (1962).

3-6. D. F. Nelson and R. J. Collins, J. Appl. Phys. 32, 739 (1961).

3-7. Kisliuk and J. Walsh, Appl. Opt., 1, 45 (1962).

3-8. Javan, et al., J. Opt. Soc. Am., 52, 96 (1962).

3-9. G. L. Clark, et al., J. Opt. Soc. Am., 52, 878 (1962).

3-10. F. McClung and R. Hellwarth, Proc. IEEE, 51, 46 (1963).

3-11. T. H. Maiman, R. H. Hoskins, I. J. D'Haenens, C. K. Asawa, and V. Evtuhov, Phys. Rev. 123, 1151 (1962).

3-12. Z. J. Kiss and R. C. Duncan, Jr., Proc. Inst. Radio Engrs. 50, 1531 (1962).

3-13. L. F. Johnson and K. Nassau, Proc. Inst. Radio Engrs. 49, 1704 (1961

3-14. G. D. Boyd, R. J. Collins, S. P. S. Porto, A. Yariv, and W. A. Hargreaves, Phys. Rev. Letters 8, 269 (1962).

3-15. F. J. McClung and R. W. Hellwarth, J. Appl. Phys. 33, 828 (1962).

3-16. E. J. Woodbury, Hughes Aircraft Company, Culver City, California (Private Communication).

3-17. L. F. Johnson, J. Appl. Phys. 33, 756 (1962).

3-18. D. F. Nelson and W. F. Boyle, Appl. Opt. 1, 181 (1962).

3-19. D. J. Blattner, Goldsmith, Z. J. Kiss, F. Sterzer, J. P. Wittke, Radio Corporation of America, Interim Report No. 5, "A Research Program on the Utilization of Coherent Light," 20 October 1962, David Sarnoff Research Center, Princeton, New Jersey.

3-20. Laser Devices Exploratory Investigation, Part 2 Technical Discussion, Hughes Research Laboratories, Malibu, California.

3-21. C.H. Townes, Proceedings of the NATO-SADTC Symposium on Technical and Military Applications of Laser Techniques, April 1962.

3-22. B.J. McMurtry and A.E. Siegman, Appl. Opt. 1, 51 (1962).

3-23. R.J. Collins, D.F. Nelson, A.L. Schawlow, W. Bond, C.G.B. Garrett and W. Kaiser, Phys. Rev. Letters 5, 303 (1960).

3-24. I.D. Abella and H.Z. Cummins, J. Appl. Phys. 32, 1177 (1961).

3-25. C.L. Tang, H. Statz, and G. de Mars, J. Appl. Phys. (inprint); H. Statz, C.L. Tang and G. de Mars, Bull. Am. Phys. Soc., Ser. II, 8, 87 (1963). A.M. Prakhorov, Third Intern. Conf. Quantum Electron., Paris, February 11-15, 1963.

3-26. C.L. Tang, H. Statz, and G. de Mars, "Regular Spiking and Single-Mode Operation of Ruby Laser," Appl. Phys. Lett., 2, 222 (1963).

3-27. P. Walsh and Kemeny, "Laser Operation without Spikes in a Ruby Ring," J. Appl. Phys. 34 (Part 1), 956 (1963).

3-28. Dieter Ross, "Torroidal Ruby Laser," Proc. IEEE (1963).

3-29. B.J. McMurtry, and A.E. Siegman, "Photomixing Experiments with a Ruby Optical Maser and a Traveling-wave Microwave Photo Tube," Appl. Opt., (Supplement 1: Optical Masers), 133 (1962).

3-30. H. Statz, G. de Mars, "Quantum Electronics," Columbia University Press, New York 1960).

3-31. W. Kaiser and H. Lessing, "Effects of an Electric Field on the Laser Emission of Ruby," Appl. Phys. Lett. 2, 207 (1963).

3-32. "Study on Optical Communications from Deep Space," Fourth Interim Progress Report, 27 March 1963 through 31 May 1963.

3-33. F.J. McClung and R.W. Hellwarth, "Giant Optical Pulsations from a Ruby," J. Appl. Phys. 33, 828 (1962).

3-34. H.C. Nedderman, Y.C. Kinag, F.C. Unterleitner, "Control of Ruby Laser Oscillation by an Inhomogenous Magnetic Field," Proc. IRE, 50, 1687 (1962).

3-35. Elian Snitzer, "Neodymium Glass Laser," Third Symposium Quantum Electronics, Paris 1963 (to be published).

3-36. G.D. Boyd, R.J. Collins, S.P.S. Porto, and A Yariv, "Excitation, Relaxation, and Continuous Maser Action in the 2.613 Micron Transition of CaF_2: u^{3T}," Phys. Rev. Letters, 8, 169 (1962).

3-37. L. F. Johnson, G. D. Boyd, K. Nassau and R. R. Soden "Continuous Operations of the Nd^{3+} $CaWO_4$ Optical Masers," Proc. IRE, 50, 213 (1962).

3-38. L. F. Johnson, "Maser Characteristics of Rare Earth Ions," J. Appl. Phys. 34, 897 (1963).

3-39. L. F. Johnson, "Continuous Operation of the CaF_2: Dy^{2+} Optical Maser," Proc. IRE, 50, 1691 (1962).

3-40. Z. J. Kiss and Duncan, Proc. IRE 50, 1531 (1962).

3-41. L. F. Johnson, "Continuous Operation of the CaF_2: Dy^{2+} Optical Maser," Proc. IRE, 50, 1691 (1962).

3-42. A. Yariv, "Continuous Operation of a CaF_2: Dy^{2+} Optical Maser," Proc. IRE, 50, 1699 (1962).

3-43. R. N. Hall, G. E. Fenner, J. D. Kingsley, T. J. Sottys, and R. O. Carlson, "Coherent Light Emission from GaAs Junctions," Phys., Rev. Letts., 9, 366 (1962).

3-44. R. J. Keyes and T. M. Quist, Proc. IRE 50, 1822 (1962).

3-45. M. I. Nathan, W. P. Dumke, G. Burns, F. H. Dill, and G. Lasher, "Stimulated Emission of Radiation from GaAs Junction, Appl. Phys. Letts. 1, 62, (1962).

3-46. I. Melngailis, "Maser Action in InAs Diodes," Appl. Phys. Letts., 2, 176 (1962).

3-47. K. Weiser and R. S. Levitt, "Stimulated Right Emission from Indium Phosphide," Appl. Phys. Letts., 1, 178 (1963).

3-48. N. Holenyak and S. F. Bevaoqua, Appl. Phys. Letts., 1, 82 (1962).

3-49. A. E. Michel, E. J. Walker, M. I. Nathan, "Determination of the Active Region in Light-Emitting GaAs Diodes," IBM Journal, January 1963, p. 70.

3-50. G. Cheroff, F. Stern and S. Triebwasser, "Quantum Efficiency of GaAs Injection Lasers," Appl. Phys. Letters, 2, 173 (1963).

3-51. I. Melngailis, and R. J. Rediker, "Magnetically Tunable CW InAs Diode Maser," Appl. Phys. Letters, 2, 202 (1962).
T. M. Quist, R. J. Keyes, W. E. Krog, B. Lax, A. L. McWharter, R. H. Ridiker, and H. J. Zeiger, "Semiconducting Maser of GaAs," Appl. Phys. Letters 1, 91 (1962).

3-52. W. E. Howard, F. F. Fang, F. H. Dill, Jr., and M. I. Nathan, "CW Operation of a GaAs Injection Laser," IBM Journal, January 1963, p. 74.

3-53. G. Burns, R. A. Laff, S. E. Blum. F. H. Sill, Jr., M. I. Nathan, "Directionality Effects of GaAs Light Emitting Diodes," Parts I and II IBM Journal, January 1963, p. 62.

3-54. M. J. Stevenson, J. D. Axe, R. J. Lankard, "Line Widths and Pressure Shifts in Mode Structure of Stimulated Emission from GaAs Junctions," IBM Journal, April 1963, p. 155.

3-55. G. Burns, M. I. Nathan, "Line Shape in GaAs Injection Lasers," Proc. IEEE, March 1963, p. 472.

3-56. G. Burns, F. H. Dill, Jr., M. I. Nathan, "The Effect of Temperature on the Properties of GaAs Laser," Proc. IEE, 51, 947 (1963).

3-57. Lasher, "Threshold Relations and Diffraction Loss for Injection Lasers," IBM Journal, January 1963, p. 58.

3-58. W. R. Bennett, Jr., Appl. Opt. Supplement 1 (1962).

3-59. A. Javan, W. R. Bennett, Jr., and D. R. Herriott, Phys. Rev. Letters 6, 106 (1961).

3-60. A. Javan, E. A. Ballik, and W. L. Bond, J. Am. Opt. Soc. 52, 96 (1962).

4. STATISTICAL STUDY OF LASER TRANSMITTERS

INTRODUCTION

Radiation generated by lasers is subject to unintended fluctuations. These are random variations of intensity, frequency, polarization, and the direction of propagation. When radiation originating in the laser is used as the carrier of information the unintended, random variations of the radiation characteristics combine with the deliberately introduced modulation. The receiver responds to the combination; therefore, the fluctuations of the transmitter appear as signal distortion, or noise.

The various types of lasers differ greatly in their statistical fluctuations. The three-level ruby laser is by far the least stable; the four-level solid lasers occupy an intermediate position; semiconductor lasers can probably be operated quite stably; and gas lasers are extremely stable. The stability of a given laser depends on the constancy of the temperature and on the power level involved. Most stable operation is obtained at the lowest power levels.

Observed fluctuations in the laser output are consequences of fluctuations of the excitation of different oscillatory modes. The complications of the fluctuation problem arise from the fact that the laser is a multimode oscillator which is seldom operated in a single mode; furthermore, relatively minor changes in external environmental conditions may cause major shifts in mode excitation. It is desirable to introduce the terminology of mode theory and to list its principal results before entering into the discussion of the experimental material on fluctuations.

Mode theory deals with the distribution of the electromagnetic field within the laser. This distribution may be regarded in the framework of different approximations. The simplest form of mode theory involves the use of uniform plane waves of finite extent. According to this point of view, given two parallel plane mirrors of finite extent, the field between them may be regarded as the superposition of plane waves traveling back and forth. The plane waves which travel longitudinally, that is, perpendicular to the mirrors, form a standing wave pattern which leads to reinforcement when the distance of the mirrors is an integral multiple of the half-wavelength.

Reinforcement takes place when $n\lambda' = 2L$, where n is an integer, λ' is the wavelength in the laser material, and L is the distance between the mirrors.* Each value of n corresponds to a frequency at which oscillations may occur provided sufficient amplification is available at that frequency. A fixed value of n characterizes a resonance or a mode. Strictly speaking, two modes of radiation must be considered for a fixed n to allow for variation in the direction of polarization. Consecutive resonances, or modes, are separated by the wavenumber

$$\Delta\left(\frac{1}{\lambda}\right) = \frac{1}{2L\eta} \tag{4-1}$$

Generally a number of modes of this type lie within the atomic linewidth.

The above plane wave theory correctly predicts the frequency separation of the principal resonances. However, other resonances may also occur. Moreover, the assumption of uniform, plane waves of finite extent is inconsistent with electromagnetic theory. A more adequate theory regards the solid laser as a dielectric box whose resonances are computed according to the known precepts of microwave theory (Reference 4-1). It is found that the resonant frequencies of cylindrical structures are determined by equations of the form

$$\chi_{lm}^2 + \left(\frac{\pi n}{L}\right)^2 = k^2 \tag{4-2}$$

where $k^2 = 4\pi^2\nu^2\eta^2/c^2$, L is the length of the cylinder, and the numbers χ_{lm} are determined from equations involving Bessel functions in χa, where a is the radius of the cylinder. The mode is characterized by the triplet l, m, n and some information concerning polarization.

The correct modes of the gas laser are found by the method of self-reproducing amplitude and phase distributions introduced by Fox and Li (Reference 4-2). Here, as in the case of solids, a triplet of integers is obtained as mode indices; the first two are related to the transverse distribution of the electromagnetic field, and the third is the number of nodal planes contained between the two reflectors.

For reasons which are explained in Reference 4-3, only axial and nearly axial modes are of interest in connection with lasers. These are the modes for which l and m are very small and n is very large. Axial and nearly axial modes are still very large in number and many may be excited within the same atomic linewidth. In good approximation the coupling between the oscillations in different modes may be neglected. However, all oscillations derive their excitation from the same pool of atomic systems and all

*The symbol λ is used to designate the wavelength in vacuum. Thus $\lambda = c/\nu$, while $\lambda' = \lambda/\eta$, where η is the refractive index of the material.

dissipate energy partly through incomplete reflection and partly through diffraction or escape of radiation to the sides that may occur when the modes are not axial. Axial modes have the least dissipation, the highest Q.

As the excitation in the material progresses to the point at which the population inversion is large enough to cause a net stimulated emission, energy begins to rise in the various resonator modes. The driving force acting on these modes is the largest for frequencies near the center of the atomic resonance, ν_a. The mathematical theory of this process in a multimode cavity has been developed by Wagner and Birnbaum (Reference 4-4), who calculated the frequency distribution of the output in terms of the atomic linewidth and the degree of excitation. Sidestepping the intricate mathematical analysis, it may be stated qualitatively that modes with more energy than others tend to grow faster and their growth rate increases with increasing Q. All feed on the same supply of excited atoms; therefore, most of the energy will be concentrated in oscillations with the highest Q and with frequencies nearest to the center frequency of the atomic line.

The Q of each mode is determined by the sum of all losses from that mode. The loss caused by incomplete reflection does not vary from mode to mode; it is determined by the nature of the reflecting layer. The diffraction loss is variable. The large loss of off-axis modes usually lowers their Q's to the point at which their excitation becomes negligible and the laser may be correctly described entirely in terms of the axial modes. The losses of the axial (TEM_{oon}) modes are about equal. Ordinarily quite a few of them lie so close to the top of the atomic line that they are excited simultaneously. However, even a small difference in the diffraction losses among modes of different types may provide a great discrimination in the excitation of these modes, and if the laser is operated sufficiently close to threshold oscillations will only occur in modes of lowest loss. A small change in the laser parameters, e.g., the distance of the reflectors, is capable of causing a large change in the distribution of energy among modes.

In the following paragraphs fluctuations of the output of the following laser types is discussed: the flash-excited ruby in free running and in the giant pulse form, the neodymium laser, and the helium neon laser.

FLUCTUATIONS IN RUBY

The output of a freely running ruby laser consists of a succession of irregular, uneven spikes of approximately 0.5 microsecond duration. When the temperature of the ruby is lowered, some of the irregularity disappears but the pulsating character of the output remains. Unless special precautions are taken, the output of the ruby laser consists of oscillations in many modes and the different oscillations attain their peaks at different times.

The presence of a series of axial modes in one ruby flash was demonstrated spectroscopically by the use of Fabry-Perot etalons (References 4-5 and 4-6). Depending on the degree of excitation, ten or more lines were

observed separated in wavenumber according to Equation 4-1. The outputs of the different modes may be mixed in a photodetector and the difference frequency detected. The spectroscopic analysis was confirmed by frequency analysis, and beat frequencies between axial and some transverse waves were found (References 4-7 and 4-8). For a ruby rod 5.62 cm long, such as used in the experiments of McMurtry, the axial modes are 1.5 Gc apart. The individual axial modes are relatively fine; their spectral width varies with temperature and excitation. McMurtry observed linewidths from 3 to 20 mc. However, the entire spectrum excited during a flash will extend over a frequency region covering over 10 Gc. The coverage is of course in the form of a fine-toothed comb spreading over a certain region — at least this is the situation during a short interval, say 1 microsecond, during the flash. Some time later a fine-toothed comb is again present but it has moved compared to its earlier position. Such a motion of the entire spectrum of the ruby is not demonstrated by mixing experiments because the difference frequencies remain the same. The shifting from one spike to the next was demonstrated by a fast time resolution interferometric technique. Hughes (References 4-9 and 4-10), as well as Hanes and Stoicheff (Reference 4-11) demonstrated conclusively that the frequencies of oscillation change from one spike to the next. A drift is observed within the same mode pattern toward lower frequencies. This amounts to about 10 mc per microsecond and may be explained by the warming of the ruby. In addition, there are sudden shifts from one mode pattern to another as the population inversion becomes exhausted in some regions of the crystal while it remains high in others, and the mode pattern readjusts itself so that the energy stored in the less frequently used regions is tapped.

Further evidence of the variability of the ruby mode pattern comes from the work of Evtuhov and Neeland (Reference 4-12) who observed the spots appearing on the end face of the ruby and correlated their pattern with the radiation pattern in the far field.

In conclusion it can be said that the excitation of the modes is strongly dependent on the power level at which the ruby is flashed. Increasing the pump energy 10 percent above threshold provides a large number of axial modes. Under such circumstances the ruby laser delivers a burst in which the intensity is highly variable and whose frequency range encompasses at least 15 Gc. The polarization of the emitted radiation will also be variable unless special precautions are taken to stabilize it.

The output of the giant pulse ruby laser is a single pulse of 0.1 to 0.5 microsecond duration. Its polarization is well defined by the optical shutter. Although details of the spectra are not available, it appears from the measurements of McClung (Reference 4-13) that the spectral range of a fast-switched giant pulse laser is about five times as wide as the spectral range of the same laser when oscillating freely, i.e., without Q-spoiling. The beam pattern of the giant pulse laser is similarly about five times broader, indicating the presence of a larger number of nonaxial modes.

Baker and Peters (Reference 4-14) demonstrated that transverse modes may be suppressed and the directivity of the laser may be enhanced. Their apparatus consists of a ruby with one external reflector with a mode-selecting aperture at the common focus of two convex lenses placed between the ruby and the reflector. The aperture eliminates radiation in transverse modes since the latter are focused off the optical axis.

FLUCTUATIONS IN NEODYMIUM LASERS

Stimulated emission from the trivalent neodymium ion has been observed in a number of ionic crystals and also in glass. The state of art as far as it pertains to Nd in crystals has been reviewed by Johnson (Reference 4-15) whose data are summarized here.

Tungstates, molybdates, and fluorides of alkali earths serve as host crystals of neodymium, the most suitable crystal being $CaWO_4$ (Scheelite). The relevant transition takes place from the $^4F_{3/2}$ to the $^4I_{11/2}$ level. These are actually multiple levels which split in the crystal field. Consequently a variety of lines is observed all in the vicinity of 1.06 micron. In $CaWO_4$, for example, stimulated emission has been observed at the following wavelengths:

At 77° K: 1.066, 1.0650, 1.0641, 1.0633, 1.0576.

At 295° K: 1.0652, 1.0582.

Thresholds for the different lines are different; the room temperature lines correspond to the second and the last of the low-temperature lines. In different crystals the wavelengths are slightly different; they depend on the charge-compensating mechanism attendant to the presence of Nd^{3+} at a divalent lattice site.

Neodymium certainly presents the physicist with a variety of lines to choose from; it also creates a problem of discrimination since at high level of operation generally more than one line will be excited. Within each line there is again a finer mode structure. The appearance of the modes is influenced by the same factors which determine the similar phenomenon in ruby.

Nd in $CaWO_4$ has been operated continuously at 85° K. CW neodymium lasers with flat end plates operate with severe spiking. The spikes account for over 90 percent of the output. With a confocal geometry, spikes are present to a limited extent (amounting to less than 5 percent of total output) at power levels slightly above threshold, but essentially no spikes are observed at power levels well above threshold. The highest power level observed for confocal rods of $CaWO_4:Nd^{3+}$ in CW operation is 0.5 watt.

Laser action of Nd ions embedded in glass was reported by Snitzer (Reference 4-16) and Young (Reference 4-17). Spectroscopic data were not

published in great detail and data on the time variation of the output intensity do not seem to be available either. It appears, however, that spikes are present in the output and that the spectral range of the output is at least as wide as in the case of Nd in ionic crystals.

FLUCTUATIONS IN NEON LASERS

Neon lasers are of interest in three spectral regions: in the visible at 6328 Å, in the near-infrared around 1.15 microns, and in the farther infrared at 3.39 microns. All neon lasers operate continuously. The intensity of the output is reported as constant, but no data were published on intensity fluctuations. It is to be assumed that spontaneous fluctuations of the intensity are not sufficiently large to interfere noticeably with the transmission of signals by means of amplitude modulation.

Multimode operation is common in neon lasers, but there is no indication of mode instability, i.e., the switching back and forth between modes. Conceivably such instability could be produced, but normally it does not seem to be present.

The visible neon laser operates within a single atomic line at 6328 Å. The power output varies with the length of the tube and with the exciting RF current. For a tube 1 meter long output powers between 1 and 2 milliwatts were reported, while 2-meter tubes gave 5 to 11 milliwatts. The doppler linewidth $\Delta\nu_D$, which determines the maximum frequency range over which oscillations may occur, is approximately 1.5 Gc. The occupied portion of this range Δ is determined by the equation

$$\Delta = \frac{\Delta\nu_D \ell n(L/L_t)}{\sqrt{\ell n 2}} \qquad (4\text{-}3)$$

where L is the length of the laser and L_t is that length at which threshold is reached (Reference 4-18). Longer tubes deliver larger output, but will do so over a broader frequency band. It is possible to select a single mode oscillation by the mode suppression device of Kogelnik and Patel (Reference 4-19).

No quantitative data were found on the frequency stability of the individual modes of the 6328 Å neon laser. It is highly probable though that the situation is not very different from that observed in the case of the 1.1523-micron laser.

The frequency stability of the 1.1523-micron laser has been explored in great detail by Javan and his associates (References 4-20 and 4-21). The measurements included the generation of beat frequencies between different modes of the same laser line as well as beat frequencies between two lasers. In the latter case, extreme precautions were taken to ensure insulation from

environmental disturbances. The following quantitative data are available for operation under optimum condition:

> The doppler linewidth for 1.15 microns is about 800 mc. Longitudinal modes of a laser 1 meter long are 150 mc apart. Short time frequency stability of a single mode is about 20 cps or 8 parts in 10^{14}. Resettability of a laser after shutdown or detuning is 0.5 mc. Power level, 0.5 megawatt.

It is to be noted that 13 atomic lines of neon have been observed in stimulated emission in the wavelength range extending from 1.08 to 1.52 microns. These are listed in Bennett's comprehensive article on gas lasers (Reference 4-22). The 1.1523-micron line dominates the others. Most of these are not seen unless the dominant line is suppressed or attenuated.

The spectroscopy of the 3.39-micron line has not been extensively developed. Stimulated emission occurs at this wavelength with a very high gain. However, the construction of a workable laser has been hindered by the large losses which are difficult to avoid in a laser at this wavelength. A power output of 10 milliwatts has been reported (Reference 4-23), but no information has been published on the frequency and power stability of this laser.

REFERENCES

4-1. J. P. Carson, S. P. Mead, and S. A. Schelkunoff, "Hyper-Frequency Wave Guides," Bell Syst. Tech. J. 15, 310 (1936).

4-2. A. G. Fox and T. Li, "Resonant Modes in a Maser Interferometer," Bell Syst. Tech. J. 40, 453 (1961).

4-3. B. A. Lengyel, Lasers, John Wiley & Sons, New York, 1962.

4-4. W. G. Wagner and G. Birnbaum, "Theory of Quantum Oscillators in a Multimode Cavity," J. Appl. Phys. 32, 1185-1193 (1961).

4-5. I. D. Abella and C. H. Townes, "Mode Characteristics and Coherence in Optical Masers," Nature 192, 957 (1961).

4-6. M. Ciftan, A. Krutchkoff, and S. Koozekanaani, "On the Resonant Frequency Modes of Ruby Optical Masers," Proc. IRE 50, 84 (1962).

4-7. B. J. McMurtry, "Investigation of Ruby-Optical-Maser Characteristics Using Microwave Phototubes," Tech. Rept. 177-3, Stanford Electronics Laboratories, Stanford University, 1962.

4-8. M. Di Domenico et al., "Optical Frequency Mixing in Bulk Semiconductors," Applied Physics Lett. 1, 77 (1962).

4-9. T. P. Hughes, "Time Resolved Interferometry of Ruby Laser Emission," Nature, 195, 325 (1962).

4-10. T. P. Hughes and K. M. Young, "Mode Sequences in Ruby Laser Emission," Nature, 196, 332 (1962).

4-11. G. R. Hanes and B. P. Stoicheff, "Time Dependence of the Frequency and Linewidth of the Optical Emission from a Pulsed Ruby Maser," Nature 195, 587 (1962).

4-12. V. Evtuhov and J. K. Neeland, "Observations Relating to the Transverse and Longitudinal Modes of the Ruby Laser," Applied Optics 1, 517 (1962).

4-13. F. J. McClung and R. W. Hellwarth, "Characteristics of Giant Optical Pulsations from Ruby," Proc. IEEE 51, 46 (1963).

4-14. J. A. Baker and C. W. Peters, "Mode Selection and Enhancement with a Ruby Laser," Applied Optics 1, 674 (1962).

4-15. L. F. Johnson, "Optical Maser Characteristics of Rare-Earth Ions in Crystals," J. Appl. Phys. 34, 897 (1963).

4-16. E. Snitzer, "Optical Maser Action of Nd^{3+} in a Barium Crown Glass," Phys. Rev. Lett. 7, 444 (1961).

4-17. C. G. Young, "Continuous Glass Laser," Appl. Phys. Lett. 2, 151 (1963).

4-18. A. D. White, E. I. Gordon, and J. D. Rigden, "Output Power of the 6328Å Gas Maser," Appl. Phys. Lett. 2, 91 (1963).

4-19. H. Kogelnik and C. K. N. Patel, "Mode Suppression and Single Frequency Operation in Gaseous Optical Masers," Proc. IRE 50, 2365 (1962).

4-20. A. Javan, E. A. Ballik, and W. L. Bond, "Frequency Characteristics of a CW Helium-Neon Optical Maser," J. Opt. Soc. Amer. 52, 96 (1962).

4-21. T. S. Jaseja, A. Javan, and C. H. Townes, "Frequency Stability of He-Ne Masers, etc.," Phys. Rev. Lett. 10, 165 (1963).

4-22. W. R. Bennett, "Gaseous Optical Masers," Appl. Optics Suppl. 1, 24 (1962).

4-23. A. L. Bloom, W. E. Bell, and R. C. Rempel, "Laser Operation at 3.39μ in a Helium-Neon Mixture," Appl. Optics 2, 317 (1963).

5. MODULATION SYSTEMS

INTRODUCTION

The choice of the modulation method for an optical communication system is dependent upon the theoretical capabilities of the various possible modulation techniques and the physical limitations of implementing those systems. To evaluate fully the various modulation techniques, the transmission characteristics of the channel and the signal noise effects of the system components must be specified. This information is not completely available at present although significant strides have been made toward the solution of the optical communication transmission and detection problems. It has been possible, however, to evaluate and compare the various modulation techniques for the limiting cases of noise sources. This has been done, first, for all of the major modulation techniques considering only Gaussian noise disturbances. Then, the most promising of these systems for optical deep space communication have then been analyzed for the situation in which the Gaussian background noise is negligible, and the system is limited by signal detection noise.

MODULATION TECHNIQUES

The many methods of laser modulation can be classified into three essential techniques. The first transforms a source signal waveform into a continuously variable modulation parameter. The second involves time sampling with continuous modulation parameters. The third is characterized by sampling time and allowing the source signal to take on only a discrete set of possible values. These three techniques are summarized in Table 5-1.

TYPE I MODULATION SYSTEMS

The communication systems of today utilizing Type I modulation techniques employ modulation methods which vary the parameters of a sinusoidal carrier waveform. An entirely general discussion of modulation systems would not necessarily be limited to such carrier waveforms; however, very little benefit, if any, is gained by assuming nonsinusoidal carriers.

TABLE 5-1. LASER MODULATION TECHNIQUES

	Type I	Type II	Type III
Time	Continuous	Sampled	Sampled
Modulation parameter (amplitude, frequency, phase, polarity, etc.)	Continuous	Continuous	Quantized
Examples	FM, AM	PAM, PPM	PCM

In considering the modulation methods, the condition that the carrier waveform is spectrally isolated from the modulation waveform must be satisfied. Under these conditions the Type I modulation systems can be described by a single equation,

$$g(t) = f(t) \cos\left[\omega_c t + \phi(t)\right] \qquad (5\text{-}1)$$

where $f(t)$ and $\phi(t)$ are functions of the input signal and are defined as the modulating waveforms, and ω_c is the carrier angular frequency. In the equation $f(t)$ is called the amplitude modulation of the carrier, $\phi(t)$ is called its phase modulation, and the derivative of the instantaneous phase, $\dot{\phi}(t)$, is called its frequency modulation.

In the Type I modulation systems, $f(t)$ and $\phi(t)$ are continuous functions of an input information signal $x(t)$ which is to be transmitted. The functional relationships between the information signal and the modulation parameters are shown in Table 5-2.

With optical communications it is possible to transmit information as a continuous parameter of the polarization of a laser beam. At radio frequencies, implementation problems have prevented this type of transmission, but at optical frequencies electro-optic polarization modulators are available. In terms of operational characteristics, it is expected that polarization modulation will behave like amplitude modulation.

TYPE II MODULATION SYSTEMS

In time-sampled systems a sample from a signal source is used to modulate a carrier waveform so that at the receiving end of the communication link a sampled representation of the signal source may be reconstructed. For a band-limited information signal of bandwidth B, the signal can be sampled at a rate of 2B samples per second, and then faithfully reconstructed at the receiver. In practice, sampling rates higher than the theoretical minimum are often required because most signals are not truly band-limited.

TABLE 5-2. FUNCTIONAL RELATIONSHIPS FOR CONVENTIONAL TYPE I MODULATION TECHNIQUES

Modulation Technique	f(t)	$\phi(t)$
Amplitude modulation	$A(1 + x(t))$	Constant (or nearly so)
Amplitude modulation (suppressed carrier)	$Ax(t)$	Constant (or nearly so)
Amplitude modulation (single-sideband, suppressed carrier)	$Ax(t)$ Plus $Ax'(t)$	Constant = ϕ_o Plus Constant = $\phi_o + \pi 2$
Phase modulation	A	$\alpha x(t)$
Frequency modulation	A	$\alpha \int_{\infty}^{t} x(t)\, dt$

α, A = constant

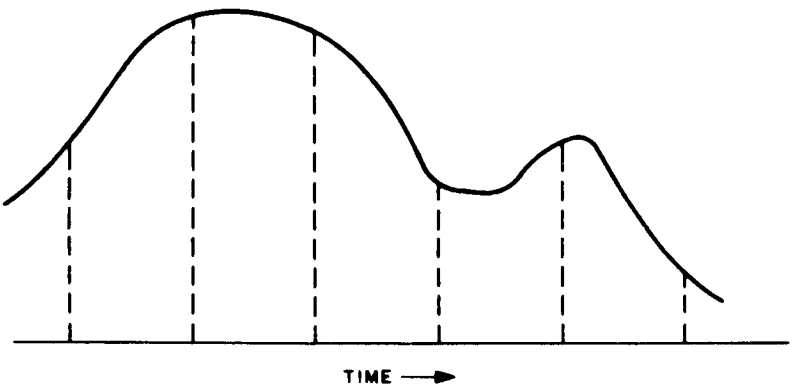
x(t) = input signal

$$x'(t) = \frac{1}{2\pi j} \int_{-j\sigma}^{+j\sigma} e^{j\omega t + \frac{\pi}{2}} \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t}\, dt\, d\omega$$

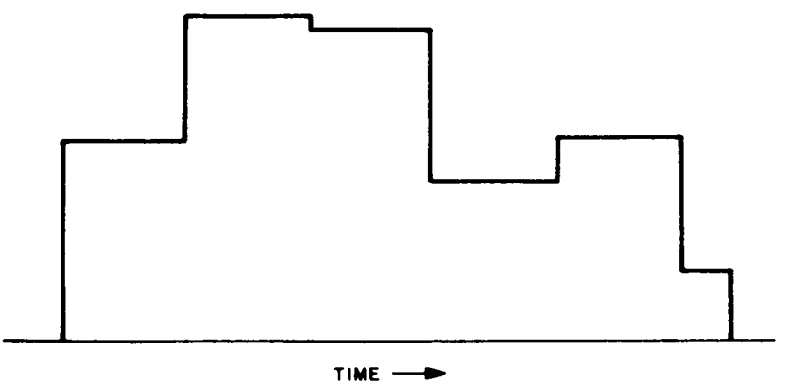
= x(t) shifted $\pi/2$ radians over entire frequency range of x(t)

A sketch of an information signal and its sampled representation is shown in Figure 5-1. Also included in the figure is a sketch of the sampled signal after it has been passed through a zero-order holding device. This latter signal could be used to modulate directly some parameter of a sine wave carrier.

With Type II modulation systems, however, it is not necessary to restrict the choice of modulation techniques to those of amplitude, frequency, and phase of the Type I systems. Waveform parameters are available for



a) Sampling of waveform



b) Zero-order hold waveform

Figure 5-1. Example of Sampled and Held Waveform

Type I systems which cannot be applied to Type II systems. These parameters include the shaping of a transmission pulse in some manner or the variation of the time occurrence of a pulse. The commonly used systems of pulse modulation are listed below.

PAM – pulse amplitude modulation

PDM – pulse duration modulation

PPM – pulse position modulation

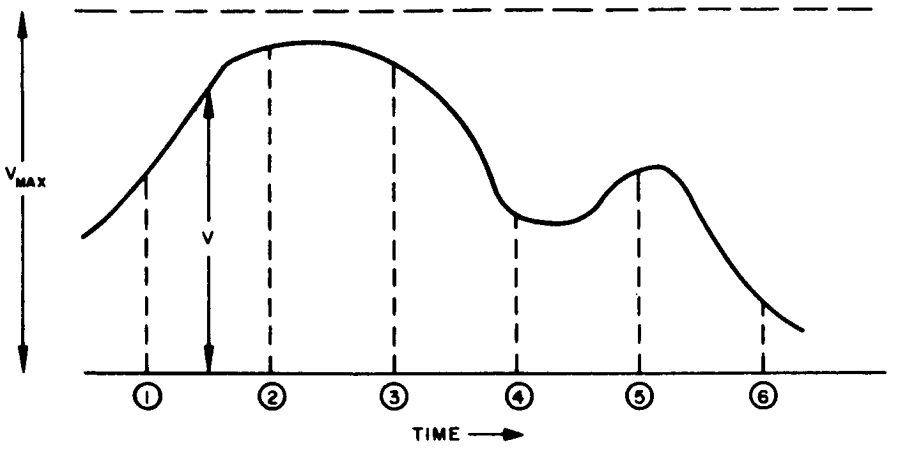
In optical communication systems a burst of the optical carrier is transmitted rather than a pulse itself. The envelope of the carrier forms a pulse, and it is the amplitude, duration, or position of this pulse envelope that carries the transmitted information.

In pulse position modulation (PPM) the signal waveform is sampled at equidistant points as shown in Figure 5-2. The signal is then coded to give transmission of short pulses of standard height whose position in the sampling time interval carries the information about the height of the sampled waveform. There are several pulse coding techniques available and the selection of a code depends on tradeoffs between bandwidth, peak signal to noise, and average transmitter power requirements. Perhaps the most straightforward coding method consists of sending a single pulse whose position (i.e., time delay) measured from the leading edge of the interval is proportional to the height of the signal waveform.

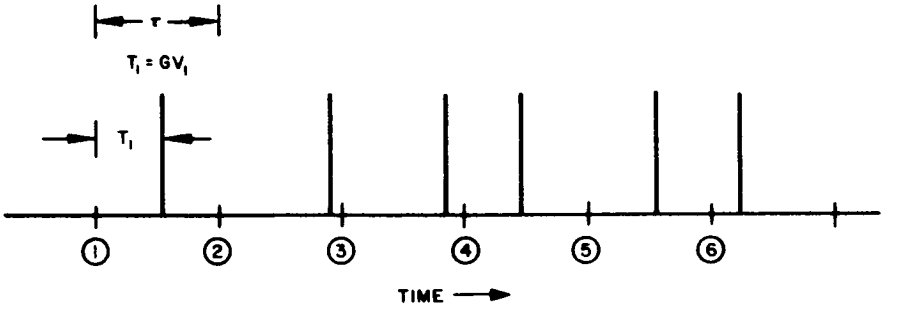
TYPE III MODULATION SYSTEMS

In Type III modulation systems the signal parameter is quantitized and the signal is time sampled. Thus, a finite number of waveforms will be used to represent each signal sample. The smallest number of waveforms is, of course, two. Modulation systems employing only two transmitter waveforms are called pulse code modulation (PCM) systems. Systems employing a larger number of waveforms have found little application in communications.

In theory the two transmitter waveforms of PCM could take any form. For optimum detection the waveforms should be the negative of each other so that matched filter detection may be employed. The usual forms of the transmitted waveforms are rectangular pulses. As with the Type II pulse modulation systems, a PCM transmission pulse is actually the envelope of the optical carrier. PCM data can be conveyed by several other means: a burst of the carrier or the absence of it, amplitude modulation, PCM/AM; an optical carrier of two possible frequencies – frequency shift keying, PCM/FSK; an optical carrier with a 0- or 180-degree phase relationship – phase shift keying, PCM/PSK; an optical carrier of right- or left-hand polarization – polarization shift keying, PCM/PL.



a) Sampling of video waveform



b) Transmitted pulse position modulation signal

Figure 5-2. Example of Pulse Position Modulation

As was mentioned for the case of Type I systems, it is expected that polarization modulation is analagous to amplitude modulation. For Type III systems, under the assumption that noise perturbation to the optical carrier and polarization state are identical, polarization modulation will be equivalent to frequency shift keying from the standpoint of communication system efficiency. In both modulation systems, two separate entities – frequency f_1 and frequency f_2 or right-hand and left-hand polarization – are detected.

It is believed that binary pulsed code modulation using opposite senses of circular polarization offers good promise for efficient transmission of information through a clear turbulent atmosphere. Limited factual support for this premise is provided by the results of recent laser transmissions over an 18-mile path between the Hughes Research Laboratories at Malibu and a Hughes field station on Baldwin Hills near Culver City. While the amplitude of the received signal scintillated over about 10 db, the linear polarization of the transmitted signal was not changed by the propagation medium. This was observed by a number of scattered measurements during the tests. The results of these limited measurements are not conclusive; however, they support what appears to be a reasonable view that atmospheric distortion probably will not lead to the spurious generation of an appreciable component of energy having circular polarization opposite in sense to that of the signal radiated at the transmitter.

The extrapolation from linear polarization measurements to circularly polarized radiation is believed to be reasonable because circular polarization is equivalent to simultaneous transmission of two linearly polarized components which are in space and time phase quadrature. There appears to be no evidence for expecting a significant amount of birefringence in the atmosphere, or for a differential effect that will occur for two orthogonally polarized components which are separated only by 1/4 wavelength at optical frequencies (i.e., circularly polarized radiation). This problem, however, must be subjected to the scrutiny of careful propagation measurements before adopting a scheme of circular polarization modulation for a fully operational space communication system capable of working through the atmosphere.

The efficiency with which data are transmitted for PCM systems is a function of both the form of modulation and the type of detection. Coherent reception in which a knowledge of the phase of a transmitter waveform is utilized during detection offers a higher efficiency than noncoherent reception systems. With phase modulation it is possible to determine the phase of a transmitter waveform from the waveform itself through means of differential coherent reception. However, the derived reference will contain noise that will degrade the efficiency of this system compared to the ordinary form of coherent reception which utilizes a relatively noise-free local oscillator for its phase reference.

In PCM communication systems, coding is often applied to the sequence of source bits in order to reduce the probability of decoding errors caused by noise corruption of the signal. Implicit error correction is

achieved by the addition of redundant symbols to the message so that in the presence of noise, the messages will differ from all other messages by a maximum amount. Additionally, coded PCM transmission is the only modulation system that offers a decrease in the energy required per information bit transmitted for arbitrarily large ratios of bandwidth-to-information rate.

THEORETICAL COMPARISON OF SYSTEMS FOR ADDITIVE, GAUSSIAN CHANNELS

To make an intelligent selection of a modulation method for an optical communication system on a theoretical basis, some manner of rating the various methods must be employed. One criterion that has been widely applied for space communications (Reference 5-1) is the maximization of the information transmission rate, H, for a given average transmission error rate, P_e. The efficiency comparison is based on a term designated as β which is defined by

$$\beta = \frac{P_M}{e^2 H} \tag{5-2}$$

where

P_M = minimum received power required per bit

H = information rate (bits per second)

e^2 = noise spectral density

The term β is related to the ratio of input signal power to input noise power, S_i/N_i, by

$$\beta = \frac{S_i}{N_i} \cdot \frac{B}{H} \tag{5-3}$$

where B = channel bandwidth. Thus, for any system β may be found from the input signal-to-noise ratio and ratio of channel bandwidth to information rate.

A lower bound for β can be derived from the channel capacity theorem of information theory:

$$H = B \lg \left(1 + \frac{S_i}{N_i}\right)^* \tag{5-4}$$

*All logarithms are to the base two.

or in terms of β_o

$$H = B\lg\left(1 + \beta_o \frac{H}{B}\right) \tag{5-5}$$

where β_o = limit of β for arbitarily small probability of error. Then,

$$\beta_o = \frac{B}{H}\left(2^{H/B-1}\right) \tag{5-6}$$

To analyze the various modulation schemes in terms of communication efficiency a description of the noise sources must be assumed. The approach to be taken here will be to assume additive white Gaussian noise for the analysis. The effects of non-Gaussian noise are considered later. A further condition to the analysis will be the assumption that analog data (e.g., TV, voice) will be transmitted rather than basically digital data. The result of this assumption will be the inclusion of a quantitization error due to the conversion from analog-to-digital form for the transmission of data by digital means.

The following discussions contain the results of a communication efficiency analysis of the basic modulation methods. The derivation of the equations may be found in Reference 5-1 and will not be repeated here.

Amplitude Modulation

If the continuous AM channel is divided into L increments, the probability of error per sample, P_e', is

$$P_e' = \frac{2(L-1)}{L} \quad \phi(-z) \tag{5-7}$$

where

$$\phi(-z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-z} e^{-x^2/2}\, dx \tag{5-8}$$

$$z = \sqrt{\frac{3}{L^2-1}\,\frac{S_o}{N_o}} \tag{5-9}$$

where

$$S_o = \frac{A^2}{12}$$

A = maximum signal excursion

$N_o = \sigma^2$ = noise power

The information rate per sample, H', for $S_o/N_o > 5$ is

$$H' = \frac{1}{2} \lg \frac{S_o}{N_o} + 0.7523\sqrt{\frac{N_o}{S_o}} - 0.2346 \qquad (5\text{-}10)$$

For a suppressed carrier, coherent detection double-sideband amplitude modulation systems the ratio of RF bandwidth B to information rate H is

$$\frac{B}{H} = \frac{1}{H'} \qquad (5\text{-}11)$$

Thus,

$$\beta_{coh} = \frac{1}{2H'} \left(\frac{S_o}{N_o} \right) \qquad (5\text{-}12)$$

For incoherent double-sideband AM systems the value of β assuming linear detection is

$$\beta_{incoh} = \frac{2}{H'} \left(\frac{S_o}{N_o} \right) \qquad (5\text{-}13)$$

for the input signal-to-noise ratio much greater than one.

A graph of β versus B/H for coherent and incoherent AM reception is shown in Figure 5-3. Markers on the curves show the value of the information rate, H.

Frequency Modulation

For frequency modulation the error rate per sample, P_e', will be identical to that for amplitude modulation:

$$P_e' = \frac{2(L-1)}{L} \phi(-z)$$

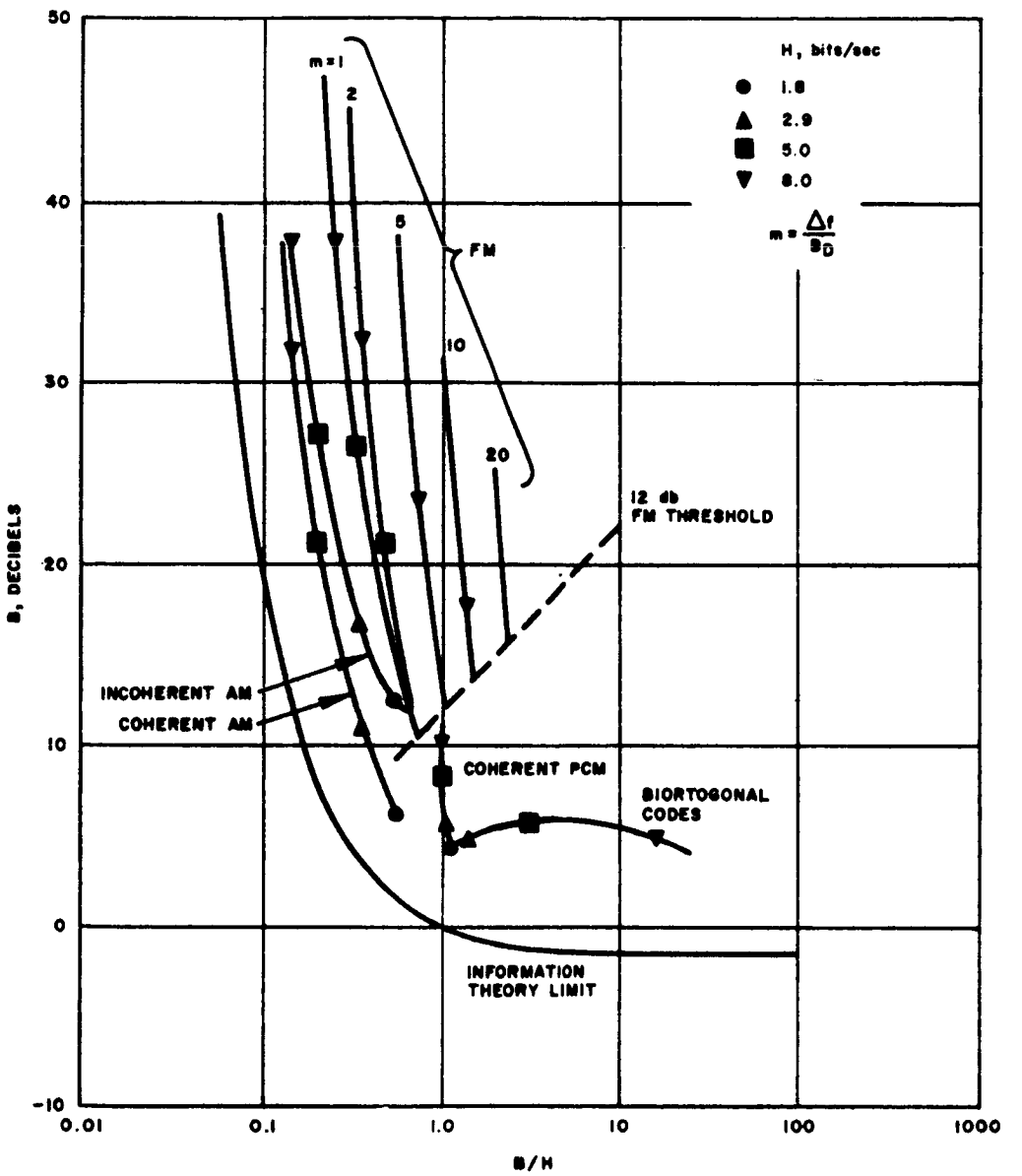


Figure 5-3. Modulation Techniques Comparison

The information rate per sample, H', for $S_o/N_o > 5$ is also of the same form for frequency modulation.

$$H' \cong \frac{1}{2} \lg \frac{S_o}{N_o} + 0.7523 \sqrt{\frac{N_o}{S_o}} - 0.2546 \qquad (5\text{-}14)$$

The value of β for frequency modulation for $S_i/N_i > 16$ is

$$\beta = \frac{1}{2H'} \left(\frac{B_D}{\Delta f}\right)^2 \left(\frac{S_o}{N_o}\right) \qquad (5\text{-}15)$$

where

B_D = data bandwidth

Δf = carrier frequency deviation

For operation at an input signal-to-noise threshold ratio of 16,

$$\beta = 16 \frac{B}{H} \qquad (5\text{-}16)$$

β is plotted against B/H for frequency modulation transmission in Figure 5-3. The communication efficiency is represented by a family of curves for various values of the modulation index (ratio of maximum frequency deviation versus data bandwidth).

Pulse Modulation

Pulse amplitude modulation (PAM) is basically the same as ordinary amplitude modulation. For AM the carrier is a single sine wave, while for PAM the sine wave and all its harmonics represent the carrier. As with AM, PAM offers no improvement in signal-to-noise ratio for an increase in signal bandwidth.

The analog of frequency modulation for time-sampled systems is the pulse position form of modulation. PPM offers basically the same improvement in signal-to-noise ratio as FM. For PPM the improvement is

$$\frac{S_o}{N_o} = \frac{1}{2}\left(\frac{t_o}{t_R}\right)^2 \frac{S_i}{N_i} \qquad (5\text{-}17)$$

where

t_o = sampling period

t_R = pulse rise time

In terms of system bandwidth,

$$\frac{S_o}{N_o} = \frac{1}{32}\left(\frac{B}{B_D}\right)^2 \frac{S_i}{N_i} \tag{5-18}$$

The relationship of β versus α will differ by at most a constant from the FM curves shown in Figure 5-3.

Pulse Code Modulation

For the PCM transmission of analog data the quantitization error due to the analog-to-digital conversion process must be considered as well as the error due to transmission disturbances. Under these conditions the error rate per sample, P_e', for equal error contributions from the two sources is

$$P_e' = \frac{1}{2L(L+1)} \tag{5-19}$$

The information rate per sample, H', is

$$H' = \lg L - \frac{1}{2L(L+1)}\left[\lg (L-1)(2L^2 + 2L - 1)\right] + \lg (2L^2 + 2L - 1) - \lg (2L^2 + 2L) \tag{5-20}$$

For a coherent detection scheme where the RF bandwidth is twice the bandwidth of the modulating signal,

$$\beta = \frac{1}{2} \frac{S_o}{N_o} \tag{5-21}$$

For a noncoherent system where the input bandwidth is twice the data bandwidth,

$$\beta = \frac{S_i}{N_i}$$

A plot of β versus α for coherent PCM is shown in Figure 5-3.

The various forms of PCM carrier modulation (e.g. PCM/AM, PCM/FSK, etc.) can be compared readily by determining the relationship of the information bit error rate to β. For a given error rate the most efficient modulation method will be that which exhibits the lowest value of β. The relative communication efficiencies of the various PCM carrier modulation techniques are shown in Figure 5-4.

Coded PCM

For the transmission of N messages in coded PCM, an equivalent of L = 2N PCM levels will be required. For square bandwidths, the bandwidth necessary to transmit the N messages f_n times per second is $B_N = f_N/N_2$. The ratio of B/H is then given by

$$\frac{B}{H} = \frac{f_N N}{2 f_N H'} = \frac{N}{2H'} \tag{5-22}$$

The error rate per sample, P_e', is

$$P_e' = \left[2N - 2\right] \phi(-Z_N) \tag{5-23}$$

where

$$Z_N = \sqrt{N} \quad \frac{S_o}{N_o} \tag{5-24}$$

The information rate per sample is

$$H' = \lg L - P_e' \lg (L-1) + P_e' \lg P_e' + (1 - P_e') \lg (1 - P_e') \tag{5-25}$$

For a maximum likelihood detector the input signal-to-noise ratio required to provide an error probability, P_e', is

$$\frac{S_i}{N_i} = \frac{2}{N} Z_N^2 \tag{5-26}$$

Then from the definition of β,

$$\beta = \frac{Z_N^2}{H'} \tag{5-27}$$

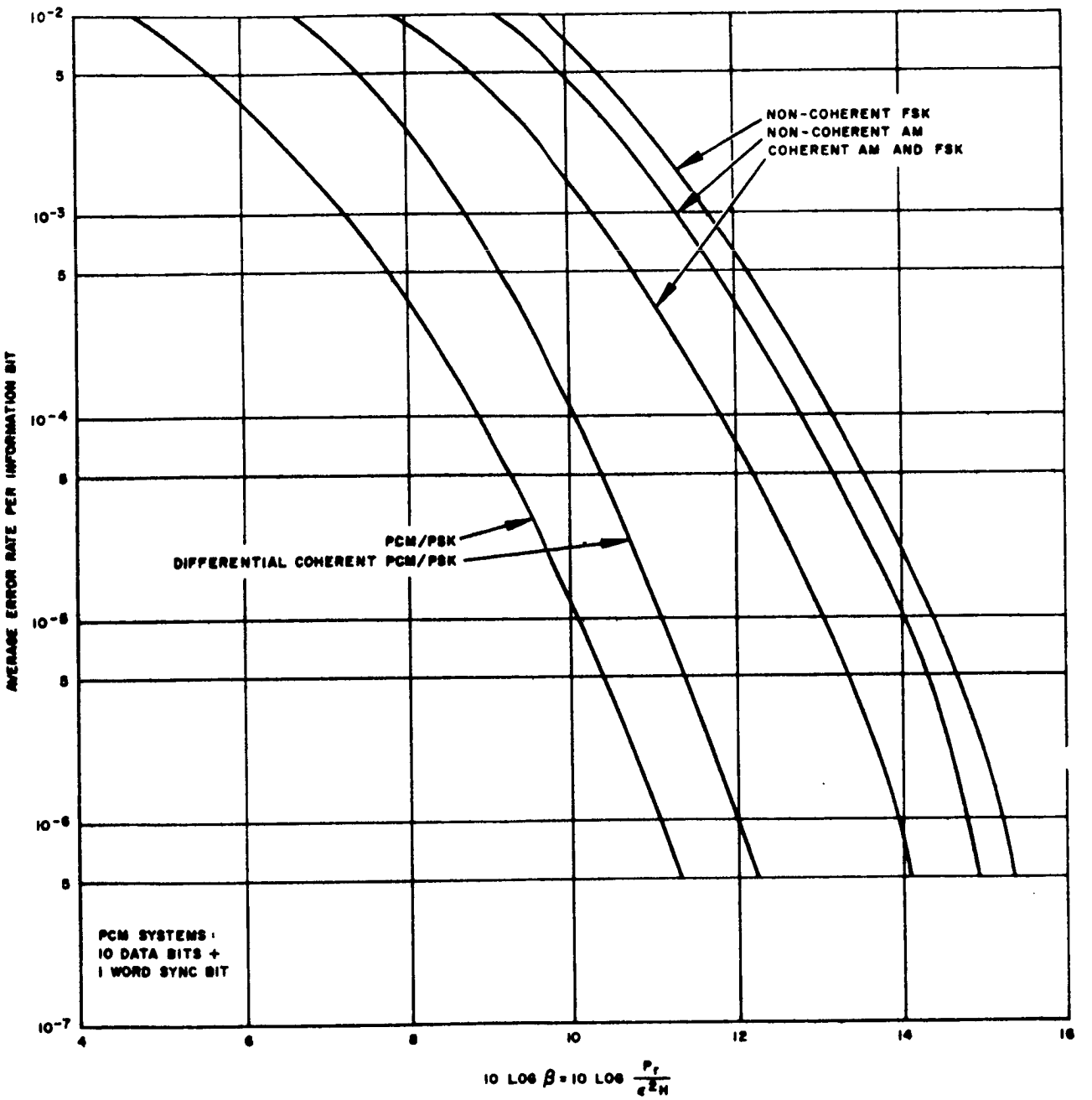


Figure 5-4. Communication Efficiences of PCM Carrier Modulation Techniques

For $P_e' << 1$

$$\beta \cong \frac{Z_N^2}{\lg 2N} \tag{5-28}$$

A plot of β versus α for biorthogonal coded PCM is also shown in Figure 5-3. With the assistance of this graph it is now possible to make a comparison of the general modulation techniques. The "best" system in a communications sense will be the one that offers the lowest value of β (minimum power required per information bit) and B/H (bandwidth-to-information-rate ratio) for a given value of information rate. In this respect the information theory limit shows that the minimum value of β will approach a limit of 0.693 for large values of B/H.

For high information rates the only system that approaches the ideal is the coded PCM form of modulation. The price of such transmission is an extremely high transmission bandwidth. For ratios of B/H equal to one the PCM transmission technique is clearly superior to all other forms of modulation. If the B/H ratio is further constrained to lower values, the FM and then the AM modulation techniques come into consideration. For a given information rate the coherent form of AM will always result in a higher communication efficiency than noncoherent AM.

From the plot of relative communication efficiencies of various PCM carrier modulation techniques in Figure 5-4 it can be seen that coherent phase shift keying is the most efficient form of modulation. This is followed by differentially coherent PSK and then by coherent amplitude modulation and frequency shift keying. The noncoherent forms of AM and FSK are next in order of efficiency.

PRELIMINARY SYSTEMS EVALUATION FOR ADDITIVE GAUSSIAN CHANNELS

Before attempting the rigorous analysis of all of the modulation techniques for nonadditive Gaussian noise effects it will be wise, because of the complexity of such an analysis, to limit the investigation to those systems which offer the greatest promise for deep space optical communications.

From the previous discussion it was shown that the coded and uncoded forms of PCM offer the highest communication efficiency of the various types of modulation methods available. For this reason the PCM studies should be continued to take into consideration the effects of nonadditive Gaussian noise. With PCM selected for further study the form of carrier modulation must be selected. Figure 5-4 illustrates the fact that coherent detection modulation systems are more efficient than incoherent systems. Unfortunately, coherent detection at optical frequencies presents many physical and theoretical problems. Optical mixing or superheterodyning will be necessary to allow

coherent detection. Investigations and experimental effort in this area are still in the very early stages. In addition, it is suggested that the atmosphere may create many difficulties for coherent reception due to variations in the phase of optical signals caused by atmospheric effects. For these reasons it appears that only noncoherent reception techniques can be profitably considered for the next phase of the analysis.

The noncoherent reception systems shown in Figure 5-4 include carrier amplitude modulation and frequency shift keying. It has been previously suggested that PCM polarization modulation is equivalent to PCM/FSK. Frequency shifting of a laser output has been demonstrated, but no practical modulation systems have been developed. Polarization modulators and amplitude modulators have been developed into practical system components.

Thus, by eliminating PPM, coherent PCM reception, and frequency shift keying as presenting serious implementation problems for the construction of an immediate experimental system, the most efficient modulation systems remaining are found to be PCM polarization modulation and PCM amplitude modulation.

Since there is very little difference between PCM/PL and PCM/AM in terms of communication system efficiency, the selection of the modulation method must be made on a total systems basis considering the physical capabilities of the systems. The transmitters of the polarization and amplitude modulation transmission systems will utilize the same laser, optical systems, and data processors. The possible physical differences in the systems will be due to the modulators and associated modulator power supplies. Presently, the most advanced amplitude and polarization modulators can be packaged to weigh only a few pounds, with the polarization modulator being slightly heavier than the AM modulator.

With radio frequency communication systems it is possible to transform a weight differential, for systems to be compared, directly into transmitted power. This is possible because the power output of a transmitter is usually dependent on the input power and not on any inherent physical limitation. However, with many of the present laser transmitters, and in particular the CW gas laser, the power output is limited by physical factors rather than pumping power. Thus, for the systems under discussion the small weight differential will not be able to be transformed into transmitter power.

An even more important consideration in system operation due to the power limitation of the laser transmitter is the form of power utilization of the modulation systems. Frequency shift keying transmits waveforms continuously, and the transmitter power which will be constant can be set at the maximum power level of the transmitter. However, PCM/AM transmits a waveform only for one of the binary symbols (i.e., binary one). Although the average transmitted power of PCM/AM and PCM/PL will be equal for an

equal distribution of transmitted ones and zeros, the transmitter must be capable of transmitting a long string of binary ones. Thus, with the same laser transmitter it will be possible to communicate with approximately twice the average power with polarization modulation than with amplitude modulation. By process of elimination, therefore, PCM polarization modulation has been selected for more extensive study.

It was mentioned previously that in an additive Gaussian channel, PPM will offer nearly the same communication efficiency as FM. Under the conditions of the so-called "signal noise" that is present only when the signal is present, it would be expected that PPM would offer a significant improvement over AM and FM. This improvement will result because of the discrete nature of PPM transmission compared to the continuous forms of AM and FM. It should be noted that since PCM is also a discrete communication system it offers this same advantage as PPM. Therefore, because of its expected advantages in an optical channel subject to signal noise, PPM has been selected for further study.

Even though the analog forms of modulation – AM, FM, and continuous polarization modulation – are not high in the efficiency ratings, they do offer the advantage of lower bandwidths, and with the exception of FM, simpler implementation. FM is not presently feasible due to the same implementation problems of frequency shifting that FSK has encountered. Continuous AM and polarization modulation components have been constructed Since it is expected that AM and PL will offer nearly the same communication efficiency, and since PCM/PL has already been selected for further study, it was profitable to investigate continuous PL for comparison with the discret form of polarization modulation.

THEORETICAL COMPARISON OF SYSTEMS FOR NONADDITIVE, NON-GAUSSIAN CHANNEL

The noise in an optical channel is caused by statistical fluctuations in the signal as well as the rms fluctuations in the external noise sources. This signal noise, of course, violates the assumption of additive uncorrelated noise sources in the previous analysis. The extent to which the conclusions of that analysis are changed remains to be investigated in an overall study that includes the combination of all noise effects.

The second limiting case of analysis in which only photon noise is considered was presented in Reference 5-2. In this report a theoretical comparison of three systems – PPM, continuous polarization modulation, and PCM polarization modulation – is made on an information theoretic basis. The information rates of the specific optical channels are computed and used for a measure of evaluation. The references to this material are given below.

For information rates of pulse position modulated optical channels, see Reference 5-2, pages 21 to 36. Information rates of continuous

polarization modulated optical channels can be found in Reference 5-2, pages 37 to 41. Information rates of PCM polarization modulated optical channels are given in Reference 5-2, pages 42 to 47.

PRELIMINARY SYSTEMS EVALUATION FOR NONADDITIVE, NON-GAUSSIAN CHANNELS

In the analysis of modulation systems considering only photon noise, the PPM and PCM systems were found to be superior to the polarization modulation form of transmission. This was also the case in the analysis of the additive Gaussian channel.

For the signal noise channel it would be desirable to have a descriptor of system efficiency such as the term β for the additive Gaussian channel. At present, such a descriptor has not been developed, and gross comparisons of the systems must be used. If only signal noise in a channel is considered, and if the pulses to be transmitted in the PPM and PCM channels are identical, the PPM system will require less transmitter power per sample than the PCM system. For a fixed sample period in which K values of amplitude resolution of the information signal are to be transmitted, PPM will require $\lg K/2$ units of energy. On the other hand, PPM will require a bandwidth of $K/\lg K$ times as great as PCM.

REFERENCES

5-1. R. W. Sanders, "Communication Efficiency of General Communication Systems," Proc. IRE (April 1960).

5-2. "Study on Optical Communications From Deep Space," Interim Progress Report, Hughes Aircraft Company, SSD 3340R, 27 March through 31 May 1963.

APPENDIX A5. NEAR-RANGE WIDE-BAND MODULATION

One potential use of the laser for communications is near-range communications from an earth-orbiting satellite. Two possible systems for this type of link have evolved: the continuous transmission channel and the burst transmission channel. In the continuous channel system, data is simply transmitted continuously, while in the burst transmission system, information is stored and transmitted over a short period of time in a low duty cycle type of operation.

Both the continuous transmission channel and the burst transmission channel attempt to utilize the large bandwidth properties of the laser to achieve a maximum information transmission rate. The continuous channel system operating in a near-range environment will not be subject to the bandwidth restrictions of deep space communication systems caused by transmitter power limitations. Thus, the same laser oscillator that would be used for the deep space system can be applied to the near-range system with a subsequent increase in information rate for the same fidelity of transmission. The burst transmission system operates at a low duty cycle and, therefore, the transmission bandwidth which is the product of the information rate and the duty factor will tend to be large.

Some of the problems associated with near-range wide-band modulation systems are discussed on the following pages. The communication system requirements in Table A5-1 have been chosen as an example for the discussion.

TABLE A5-1. SYSTEM REQUIREMENTS

Information rate	10^6 bits per second
Range	Near earth orbital
Burst rate	One burst per second
Burst duration	1 millisecond

BURST TRANSMISSION OPERATION

Burst-type communication systems have been proposed as a means of obtaining secure communications and easing the satellite tracking requirements. Requirement easing will be of great importance for extremely narrow-beam systems for which beam tracking accuracy and stability requirements are now severe. For such systems the receiver would search for and acquire the transmitter beam and then lock on to it only momentarily to achieve transmission.

For the requirements postulated, the burst transmission system will gather data over a 1-second interval, store it, and then transmit it to earth in a single 1-millisecond burst. The data may be generated by a single source (e.g., 500 x 400 element picture in a 5-bit pulse code modulation) or multiple sources (e.g., 25 4-kc voice channels in a 5-bit pulse code modulation). For the megabit information rate, the transmission rate will be in the Gigacycle range over the burst period. This high rate certainly places stringent requirements on all of the system components including the laser oscillator itself. At these high rates the analog forms of processing and modulation are quickly eliminated, and the burden placed on the task for digital components will be great.

Burst Transmission Modulation Techniques

The only discrete forms of modulation that are applicable to burst transmission are pulse code modulation (PCM) and pulse position modulation (PPM). Their advantages and disadvantages have been discussed previously and will not be repeated except for the matter of the relationship of transmission rate to information rate for the two systems.

With K resolution elements per data sample period, the transmission rate, R_{PCM}, for PCM transmission will be

$$R_{PCM} = NR_I$$

where

R_I = information rate (inverse of sampling period)

N = number of transmitted bits per sample

The term N is equal to the largest integer value of log K since an integral number of bits must be transmitted each sample period. Similarly, for PPM the transmission rate, R_{PPM}, will be

$$R_{PPM} = K R_I$$

Then,

$$\frac{R_{PPM}}{R_{PCM}} = \frac{K}{N}$$

Thus, for an equivalent sampling period, PPM will require a significantly higher transmission bandwidth. Conversely, for systems of equal transmission bandwidth, the information rate PPM will be lower than that of PCM by the same ratio of K/N as shown in Figure A5-1. This will be a strong factor in the selection of the modulation method for bandwidth-limited systems.

Information may be conveyed in a burst by an optical carrier constant over the burst duration or by a carrier consisting of a series of pulses. For the reasons stated in the previous general discussion of laser modulation techniques, the present or near future discrete forms of carrier modulation of a laser beam will be limited to PPM, PCM/AM, and PCM polarization modulation.

Burst Transmission Laser Oscillators and Modulators

The continuous wave type of laser oscillators can be made to operate in a burst mode by simple carrier chopping techniques, but at a great loss of efficiency. The present "pulsed" lasers which produce a carrier when excited by a flash lamp or pump current are much more suitable for burst communication purposes.

Solid-state lasers inherently produce spikes which may be used for information transmission. Typically, the spikes are of 1 microsecond duration with a repetition rate of up to 10^6 pps. The upper limit of the repetition rate is governed by the natural frequency of the laser system. By increased pumping power the limit can be raised somewhat, but the device efficiency drops appreciably. The solid-state laser is, therefore, nearly two orders below the postulated Gigacycle bandwidth requirements.

Semiconductor lasers emit a light pulse whenever excited by a pulse of pump current. The present repetition rate for high-power pulses for these devices is well below 10^6 pps due to heating effects when the material is pumped at high rates. This fact will restrict use of semiconductor lasers in a burst communication system unless some method of heat removal from the relatively small semiconductor junction can be developed.

Some effort has been made to smooth out the spikes of a ruby laser over the burst period to obtain a constant carrier. The results have not been too encouraging for burst communications; the resultant carrier bursts are uneven in amplitude and spectrally degraded. It may be possible, however, to utilize polarization modulation with the ruby laser. The neodymium* laser offers some promise for burst carrier transmission. Bursts of 100 kilowatts over a 1-millisecond period have been obtained with reasonably good amplitude

* N_d^{3+} in glass.

and spectral stability. However, the problems associated with repetitive operation of the laser have not been deeply explored.

The internal techniques of modulation for the solid-state lasers, while attractive from a power efficiency standpoint, are limited to transmission rates of about 1 megacycle. External methods of amplitude and polarization modulation are possible up to the Gigacycle region of transmission bandwidth.

Data Storage Limitations

Many scientific and nonscientific articles make claims, such as "The entire Bible can be transmitted in a fraction of a second on a single laser burst." In general, such claims are false not so much from deficiencies of lasers, but from the physical limitations of the peripheral equipment of a laser communication system.

If a laser burst system with 1-millisecond bursts at a 1-second repetition rate is postulated for one frame per second television (500 x 400 element picture) each burst would have to carry 10^6 bits of information in a 5-bit pulse code modulation. The modulation bandwidth for this amount of data is within the limitations of the laser. However, for this system the data would have to be collected and stored in the transmitter for a deadtime of nearly 1 second and then removed from storage and transmitted in 1 millisecond. At the receiver the collection and storage time would again be 1 millisecond. The storage rate for such an operation of 10^9 bits per second is well beyond the state of the art for storage media. Figure A5-2* illustrates the limits of capacity for feasible storage systems. The storage requirements for pulse code modulation television transmissions on a millisecond laser burst are indicated by a cross on the figure.

What is a feasible limit for the amount of information carried in a laser pulse that can be stored by a presently available storage unit? From Figure A5-2 it can be seen that the maximum storage rate per bit of information stored in the memory will be 1/0.8 microseconds for a memory storage capacity of 10^5 bits for a magnetic core memory. For a millisecond laser burst the number of information bits, N, that can be placed on the pulse will be

$$N = \frac{10^{-3} \text{ seconds}}{0.8 \times 10^{-6} \text{ second per bit}} = 1250 \text{ bits}$$

With a 50-nanosecond tunnel diode memory, the situation is somewhat improved. In this case N will be

*J. A. Rajchman, "A Survey of Computer Memories," Datamation, December 1962, pp. 26-30.

$$N = \frac{10^{-3} \text{ seconds}}{50 \times 10^{-9} \text{ second per bit}} = 20,000 \text{ bits}$$

In both cases the storage rate available for the memories falls far short of the 10^6 bits required even for one frame per second television. It may be possible to improve this situation somewhat by time multiplexing parallel storage devices or data sources and links. However, this approach soon leads to serious engineering design problems.

CONTINUOUS TRANSMISSION OPERATION

The operation of a near-range earth-orbiting communication system does not cause any theoretical problems in addition to those encountered by the deep space system. The differences between the systems will be in the components and will be caused by the larger transmission bandwidth of the near-range system.

The laser oscillators that seem most applicable for the near-range system are the gaseous laser for CW operation and semiconductor lasers for pulse and CW operation. The transmitter power requirements for near-range communication which are in the milliwatt range are achievable by many of the present-day lasers. The gas laser is capable of the megabit information rate using pulse code modulation polarization or amplitude modulation. The heating effects of the semiconductor laser, however, create problems at the megabit information rate for both CW and pulse operation.

SUMMARY

Early in the development of the laser, burst transmission schemes were proposed as a means of utilizing the "inherently large" bandwidths of lasers. After further growth of the field, it can be seen that the utilization of large optical bandwidths is not without serious problems. The picture should not be painted too black, however, for with the neodymium and the ruby laser, external modulation techniques, multiple data sources, and multiple storage devices it may be possible to approach megabit information rates over millisecond pulses. With such an information rate, voice, scientific and engineering data, and picture transmission will be possible.

Near-range transmission on a continuous basis does not appear to offer any appreciably serious problems for megabit information rate systems. The most promising system presently appears to be transmission with a gas laser, such as He-Ne and He-Xe, and solid-state laser such as $Dy^2 : CaWO_4$, and $Wd^3 : Ca\ F_2$, using pulse code modulation polarization.

---

* Nd^{3+} in glass.

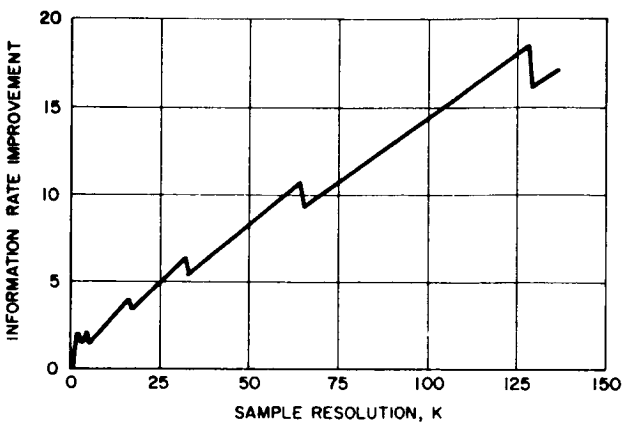


Figure A5-1. Ratio of Information Rate of PCM Versus Sample Resolution

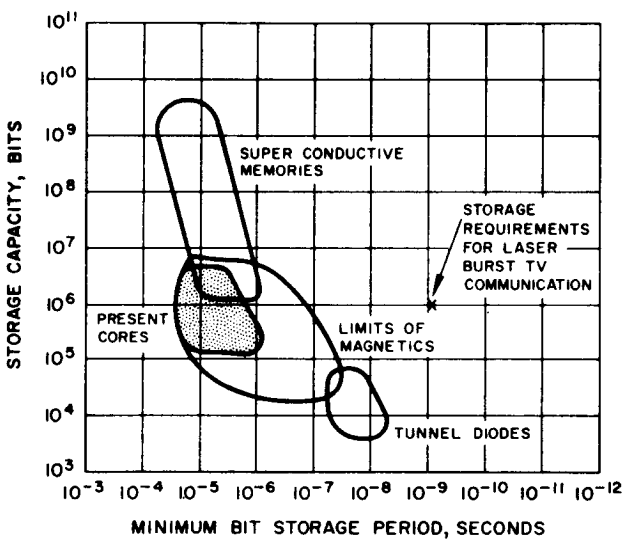


Figure A5-2. Limits of Storage Capacity

6. MODULATION DEVICES

INTRODUCTION

The following section presents a survey of the various available modulation devices to determine their ability to provide amplitude, phase, frequency, and polarity modulation for lasers suitable for deep space communications. A general discussion of modulation mechanisms precedes the preliminary evaluation of various devices.

Modulation devices may be classified as being either internal or external depending on whether the laser output is varied within the laser itself or in a separate stage following the laser oscillator.

In internal modulation the effect is enhanced by the Q of the laser cavity (ratio of stored energy in cavity to dissipated energy per radian of oscillation) so that primary power requirements are generally smaller. Moreover, there is no loss of laser signal already generated as is inevitable when the light intensity is varied external to the laser. Thus the overall system efficiency could be expected to be greater.

Most of the present day problems stem from the relative newness of the laser devices. Thus, for example, gas lasers must operate with a high Q optical cavity, and attempts to introduce a modulation element into this cavity can easily quench the oscillations altogether. In addition, investigators are still searching for means to improve the frequency and amplitude stability of the laser. Thus it is natural that the various problems be separated and more emphasis be placed on external modulation techniques as laser technology is improved. Nevertheless, two distinct disadvantages of internal modulation should be recognized. First, most of the techniques result in a non-linear response of the level of oscillation with changes on cavity parameters. More seriously, the bandwidth is limited by the high Q of the optical cavity. The path of the average photon in a gas laser illustrates the problem. Assuming 99 percent reflection by the output end mirror (and 100 percent reflection at the other end of the cavity), the average photon makes 100 round trips within the laser cavity before being emitted. If this cavity is 1.5 meters long, the photon requires 1 microsecond to escape. Thus the modulation rate must be limited to less than a megacycle. Clearly, for a given gain per unit length of the laser material, a tradeoff exists between bandwidth and output power.

The properties generally descriptive of external modulation are just the reverse of those attributable to internal modulation. Larger bandwidths and more linear responses are to be expected at the cost of low efficiency. The comparison is appropriate when a particular effect is utilized to modulate the laser beam either internally or externally. With the above stated characteristics in mind, a preliminary evaluation of these modulation methods which have been investigated to date can be made.

Portions of the following discussion have appeared in the Third Interim Report of the Deep Space Optical Communication Study and are referenced accordingly:

Third Interim Report – Pages 1-19

6.2 Intensity Modulation Mechanisms

6.2.1 Electro-Optical Effect

6.2.2 Pump Power Modulation

6.2.2.1 Semiconductor Laser

6.2.2.2 Gas Laser

6.2.2.3 Solid State Laser

6.2.3 Pressure Modulation

6.3 Polarization Modulation Mechanisms

6.4 Frequency and Phase Modulation Mechanisms

6.4.1 Electro-Optical Effect

6.4.1.1 Frequency Modulation

6.4.1.2 Phase Modulation

6.4.2 Zeeman Effect

6.4.3 Magnetostrictive Effect

ELECTRO-OPTICAL MODULATORS

Since the Third Interim Report there have been no major developments in the field of laser modulation devices. The most efficient method of external modulation at high frequencies still appears to be one which would employ the electro-optical Pockels effect of certain crystals.

Electro-Optical Materials

The relatively low electro-optical coefficients of presently known Pockels effect materials pose serious limitations on the index of modulation and the power requirements at high frequencies. Mechanisms to enhance the Pockels effect are the object of continuing investigation. The Fabry-Perot, traveling wave, and multi-element cascaded modulators are some of the mechanisms used.

Crystals which have the possibility of displaying electro-optical properties belong to either the $\bar{4}2m(V_d)$ or the $\bar{4}3m(T_d)$ crystal symmetry classes. To the first class belong the dihydrogen phosphates, so far the most successful of the crystal modulators. Deuterated potassium dihydrogen phosphate (KD*P) has the largest electro-optical coefficient of any material known today ($r = 1.7 \times 10^{-11}$ m/v). It has also the advantage of being relatively stable at room temperatures and easily grown in a laboratory. Among its disadvantages are a high dielectric constant and loss tangent (45 and 0.01 respectively) which limits its high frequency performance, the necessity of applying the electric field in the same directions as the light propagation for a maximum electro-optical effect, a narrow field of view (< 2 degrees), and a spectral transmission which extends only to the near infrared (1.2 microns). Crystals of the $\bar{4}3m(T_d)$ class do not suffer all of these disadvantages. The electric field is applied perpendicular to the direction of light propagation and the field of view is large (< 18 degrees). Copper chloride (CuCl) is such a crystal. It has the largest spectral transmission of any Pockels crystal known (0.4 to 15 microns). However, its electro-optic coefficient is only 1.8×10^{-12} cm/volt. Sphalerite (ZnS) and (ZnSe) have a somewhat larger electro-optic coefficient and a high index of refraction, however, in all cases the crystals are very hard to obtain in nature or grow in a laboratory. Gallium arsenate (GaAs) crystals have also displayed a small Pockels effect. Some organic crystals such as methenane (HMTA) and adamantane belong to the $\bar{4}3m$ class. HMTA has proved (Reference 6-1) to possess valuable properties as a light modulator. Its electro-optical constant is 7.1×10^{-12} m/v, it is stable and easily grown, and its dielectric constant and loss tangent are 6 and 0.003 respectively.

Some crystals display non-linear dielectric properties at their curie temperature. Potassium tantillumnate ($KTaO_3$), for instance, has been shown to have electro-optical properties comparable to the dihydrogen phosphates (Reference 6-2). Its curie temperature however is near 4°K.

Table 6-1 shows the electro-optic properties of the most important crystals of both classes. The voltage required for half-wave retardation has been calculated for both intensity and phase modulation as described in the earlier report. It is noted that KD*P (deuterated potassium dihydrogen phosphate) offers the lowest voltage for both applications. Sphalerite (ZnS) also has good possibilities as a modulator. Although its electro-optical coefficient is low compared to the dihydrogen phosphate, its index of refraction is much greater, a fact that becomes more significant when raised to the third and fourth powers as is the case in the applications studied. Furthermore, as pointed out earlier, cubic crystals like sphalerite can be used with

TABLE 6-1. PROPERTIES OF ELECTRO-OPTIC CRYSTALS

	KD*P	KH_2PO_4	$NH_4H_2PO_4$	RbH_2PO_4	KH_2AsO_4	ZnS	CuCl
Electro-optical constants (10^{-7}cm/kv)	r_{63} 23.6 r_{41} –	10 8.6	8.5 28	11 –	11.6 –	– 2	– 1.8
Half-wave* voltage, kilovolts	3.4	7.8	9.6	7.3	6.2	8	18
Index of refraction	–	1.509	1.524	–	1.567	2.368	1.973
Dielectric constant	45	21	15.4	15	21	8.4	
Transmission range, microns	0.4-1.2	0.4-1.2	0.4-12	–	–	–	0.4-15
Half-wave** voltage, kilovolts	2.3	5.2	6.3	4.9	4.0	3.4	9.1

*Voltage necessary to produce half-wave phase retardation when light and electric field are parallel to the optic axis (λ = 0.5 microns)

**Voltage necessary to produce half-wave phase retardation when light travels parallel to the y axis, is polarized along x axis and the E field is parallel to the z axis.

much larger apertures. Unfortunately, to date ZnS crystals are difficult to obtain. Thus for laser beams in the 0.5 to 1.2 micron spectrum the best choice of Pockels effect modulator seems to be KD*P. It is, however, by no means certain that there are not other materials, not yet developed, which might offer better properties for light modulation.

Pockels Cell Modulators

Video Light Modulators

At low frequencies most Pockels effect crystals display relatively strong piezoelectric effects. The piezoelectric strains cause additional phase

retardation. Above the piezoelectric resonant frequency this effect is lost. The resonant frequency is given approximately by the expression

$$f = \frac{48}{d} \text{ kc} \tag{6-1}$$

where d is the crystal's square edge dimensions in inches. The change in optical retardation before and after resonance is most pronounced in ADP. In KDP, which has a lower piezoelectric coupling coefficient, the effect is much smaller. A one-inch-square KDP crystal light modulator, for instance, will be resonant at 48 kc showing approximately a 10 percent change in response before and after resonance.

Since relatively large voltages are required to induce a significant optical retardation the upper frequency limit in the sub-microwave region is usually set by heating of the Pockels cell electrodes due to high RF currents. To make an approximate estimate of the currents involved, the Pockels cell may be considered a pure capacitance (the loss tangent is < 0.05 percent). If V is the maximum applied voltage, ω_m the modulation frequency, and C the capacitance, the RF current, I, that the electrodes are required to withstand is

$$I = \omega_m CV \tag{6-2}$$

The types of electrodes used may be either simple copper plates with holes to allow passage of light to gold grids or rings designed to minimize fringing of the electric field. The maximum current through the electrodes should be kept smaller than 0.5 ampere. If C is the capacitance in pico-farads, V the maximum electrode voltage in kilowatts, and f the frequency in megacycles, Equation 6-2 becomes

$$V = \frac{80}{Cf} \tag{6-3}$$

For a 4 cm^2 KD*P crystal L cm thick the capacitance is

$$C = \frac{15.9}{L} \text{ pf}$$

substituting in Equation 6-3

$$V = \frac{5L}{f} \tag{6-4}$$

This equation is plotted in Figure 6-1. Video frequencies with 100 percent modulation are then possible since in the case of KD*P half-wave retardation

is obtained with3.4 kilovolts. The Pockels effect may be enhanced by stacking a series of Pockels crystals mounted electrically in parallel and oriented in such a manner that their x - y crystal axes are at 90 degrees with respect to one another. When two such crystals are mounted as shown in Figure 6-2 only half of the voltage is necessary to produce the same effect in one crystal. The modulation voltage is reduced by a factor equal to the number of stacked crystals used. The overall capacitance of the system is of course increased by this factor. However, the required input current remains constant.

The fringing of the electric field between the electrodes provides a field in the transparent region. This introduces non-uniform phase retardations over the aperture and results in a higher voltage than that theoretically calculated to obtain a given response. These effects will not be intolerable if the modulator contrast ratio is not especially sensitive to the specific retardation.

The power dissipated in the crystal itself due to the loss tangent, or dissipation factor, can be theoretically calculated. It is

$$P = \omega_m CV^2 = \omega_m \epsilon'' \epsilon_o \frac{A}{L} V^2 \qquad (6\text{-}5)$$

where the parameters are the same as defined above and ϵ'' is the reactive complex relative permittivity. For a KD*P crystal if $A = 4\,cm^2$, $d = L\,cm$, $V = 1$ kv, $\omega_m = 2\pi \times 10^7$ cps, $\epsilon'' = 0.01$ and $\epsilon_o = 8.85 \times 10^{-14}$ fd/cm,

$$P = 2\pi \times 10^7 \times 10^{-2} \times 8.85 \times 10^{-14} \times 4 \times 10^6 = 0.222 \text{ watt}$$

a power dissipation which probably can be handled by the crystal. However, power being proportional to the square of voltage, an increase in voltage may rapidly surpass the power limit of the modulator. Equation 6-5 is plotted in Figure 6-3 for several crystal thicknesses.It is assumed that 0.5 watt is the maximum power that a crystal of the sizes being considered can dissipate. It is shown that at high frequencies 100 percent modulation will not be possible.

Perhaps the main limitation toward practical use of a Pockels effect modulator at video frequencies is the power supply. It should have the following minimum requirements:

Voltage:	1000 volts
Current:	0.1 ampere
Frequency response:	Flat to 2 mc

Such video power amplifiers have been developed for TV transmitters. The amplifier size is approximately 10 by 20 by 6 inches and would weigh 2 to 3 pounds. It requires a conventional 1000 and 300 volt dc power supply.

One possible Pockels cell intensity modulator suitable for video modulation could take the following form. To obtain 100 percent intensity modulation with less than 1 kilovolt an arrangement similar to Figure 6-2 would be used. If three such KD*P crystals were mounted using an optically biasing quarter wave plate, a theoretical peak voltage of only 0.54 kilovolt would be required to obtain 100 percent modulation. Using 2 cm cube crystals at a maximum modulation frequency of 5 mc the electrode current and power dissipation would be kept well below the limit as shown in Figures 6-1 and 6-3. The unit could be mounted within a bath of mineral oil in a quartz container which would facilitate cooling and reduce reflection losses. The light modulator size would be about 2.5 by 2.5 by 10 cm and would weigh about 2/5 pound.

Microwave Light Modulators

Light can be modulated at microwave frequencies by using the electro-optical effect. Two general approaches may be taken to do this. One consists in building a high-Q structure which would apply large field strength across the crystal with relatively low power inputs. A high Q re-entrant type, capacitively loaded coaxial-line resonator has been used to build such a modulator at GC frequencies and few watts input. The high Q requirements of such a modulator make the device inherently narrow band. A second approach consists of making a traveling electric field interact with light propagating through the electro-optical medium. If the velocity of the microwave traveling electric field could be synchronized with that of the light beam the Pockels effect would be enhanced resulting in the possibility of wide-band light modulation and economical use of microwave power. Both approaches will be studied below.

Resonant Cavity Modulator. A cross section of a cavity resonator is shown in Figure 6-4. Its resonant frequency can be determined with reasonable accuracy by setting the susceptance of the capacitance formed by the crystal equal to the inductive reactance of the coaxial transmission line formed by the outer and center conductors. Due to the necessity of boring a hole through the center of the cavity to pass a collimated light beam through, design equations available for the conventional re-entrant cavity can only approximate the actual characteristic of the particular physical configuration employed. The cavity's resonant wavelength, shunt impedance, and unloaded Q have been derived (Reference 6-3). The RF losses in the crystal must be taken into account by paralleling the shunt resistance with the dielectric resistance of the crystal R_c where

$$R_c = \frac{1}{\omega_m C \tan \delta} \tag{6-6}$$

tan δ being the loss tangent. The power dissipated in the crystal is

$$p = \omega_m C V^2 \tan \delta = \omega_m \epsilon'' \epsilon_o \frac{A}{L} V^2 \tag{6-7}$$

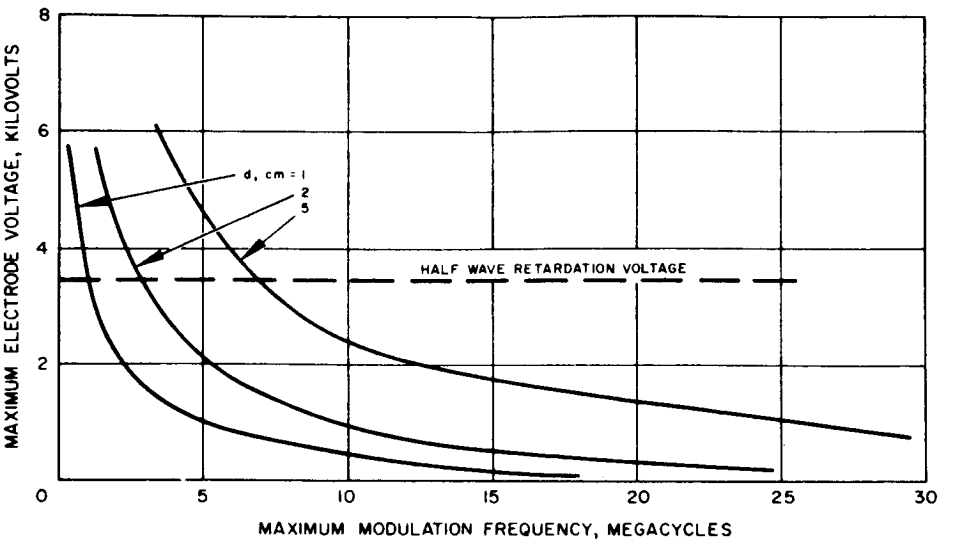


Figure 6-1. Maximum Electrode Voltage at Different Frequencies of KD* P Pockels Effect

Electrode current limited to 0.5 ampere

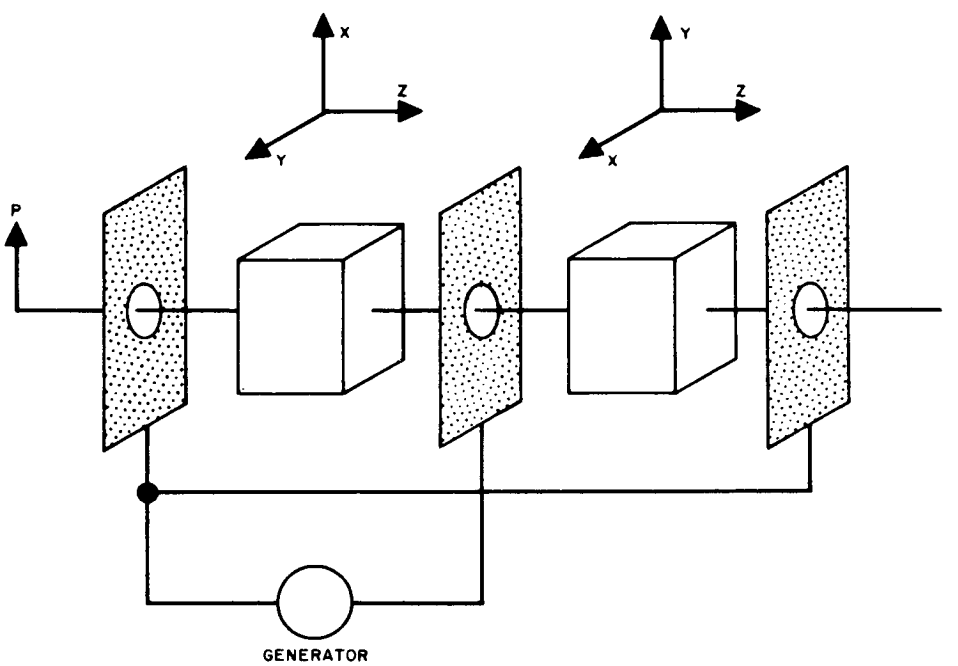


Figure 6-2. Two-Stage Pockels Effect Light Modulator

Power dissipation is limited to 0.5 watt

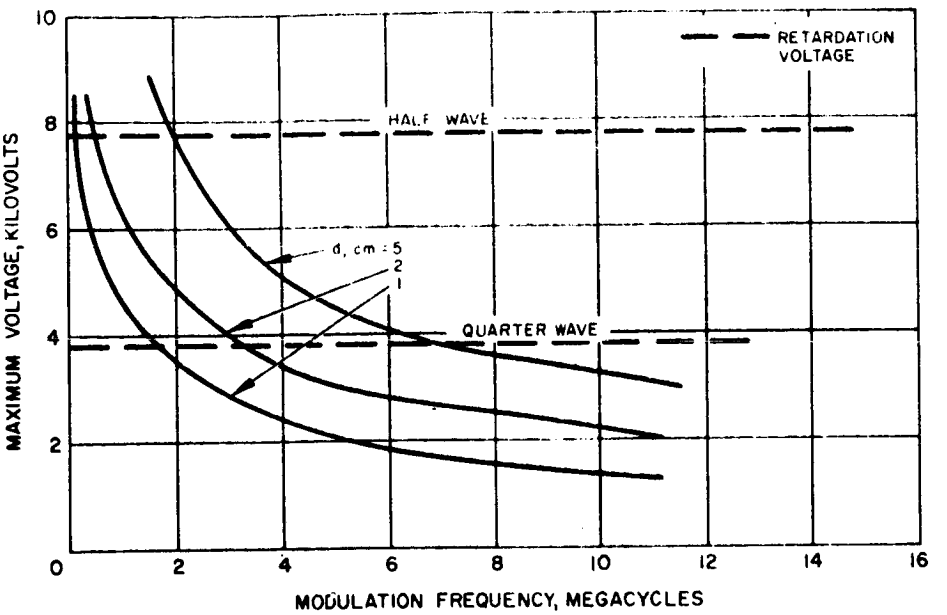


Figure 6-3. Maximum Electrode Voltage at Different Frequencies of KD*P Crystal

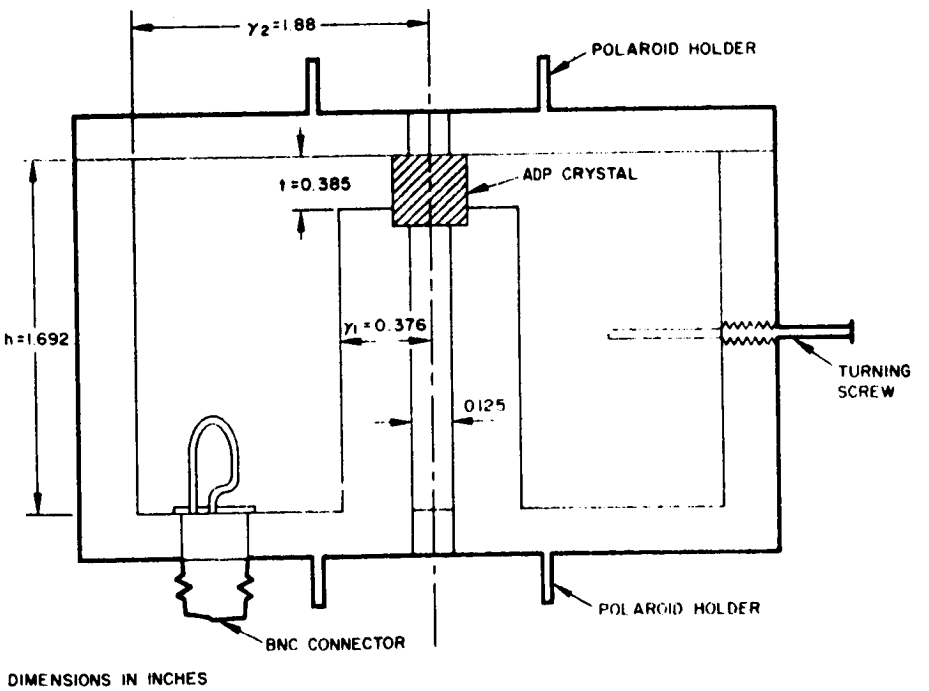


Figure 6-4. Cross section of Reentrant-Type Cavity Used for Microwave Light Modulator

A cavity resonant at 1150 mc having a Q = 1300 and tunable between ±15 mc (using a tuning screw) has been designed. Dimensions are shown in Figure 6-4. The electro-optical crystal used is KDP. Figure 6-5 shows a plot of power input to the cavity and dissipated by the crystal as a function of percent intensity modulation. The high Q of the cavity does not allow broadband modulation. For this particular cavity the bandwidth is

$$B = \frac{f_m}{Q} = \frac{1150 \times 10^6}{1300} = 0.9 \text{ mc} \qquad (6\text{-}8)$$

An increase in bandwidth can be obtained by lowering the Q of the cavity and increasing the power input.

The size of the cavity is approximately 30 cubic inches with a weight of less than 5 pounds. Equipment to produce 10 watts at 1000 mc, including power supply and temperature control devices, would weigh about 100 pounds.

Traveling Wave Modulator. Wideband light modulation can be provided by traveling wave interaction in the electro-optic crystal. The Pockels effect is enhanced by making the polarized light beam travel in synchronization with the forward component of the microwave field. In a KDP crystal the speed of light is normally three times that of microwaves. To synchronize the speed, a lightly loaded cylindrical cavity containing a Pockels crystal rod of small diameter and excited in the TM_{013} mode may be used. However, the fact that most of the microwave energy is dissipated in the cavity and therefore most of the stored field is located outside the rod makes this structure quite inefficient. Figure 6-6 shows a traveling wave modulator of this type. The length of the rod is an integral number of half wavelengths. The longer the length the stronger the electro-optical interaction. When the velocities are matched, the forward wave provides a constant electric field at a given point along the light wave train, while in the backward wave the phase retardation cancels out.

Using a KDP crystal rod 0.4 cm^2 by 3.4 cm long, a bandwidth of 5 mc centered at 2700 mc was obtained (Reference 6-4). Figure 6-7 shows the microwave power needed as a function of percent intensity modulation. With 10 watts input power, over 40 percent intensity modulation can be achieved. The weight of the modulator would not exceed 5 pounds. However, estimated weight of associated equipment including power supplies to produce 10 watts at 2700 mc would run above 200 pounds.

The efficiency of the traveling wave modulator described can be improved if the TEM wave would have most of the electric field in the electro-optical crystal. In such a case the microwave structure could be made essentially dispersionless, a necessary requisite for a wideband light modulator.

It can be shown (Reference 6-5) that the conditions for synchronization can be satisfied when the modulating field is propagated as a TEM wave along

a plane-parallel guide. The optical z-axis of the crystal is oriented perpendicular to the parallel planes. A traveling wave phase modulator can be built using the principles previously described. The cross-sectional view of a phase modulator is shown in Figure 6-8. The parameters are calculated from the following relationships (Reference 6-6).

$$a = \frac{w(e_1 - 1)}{n^2 - 1} \tag{6-9}$$

and

$$b = Z_o \frac{\left[a(a - w + e_1 w)\right]^{1/2}}{120\,\pi} \tag{6-10}$$

where

e_1 = relative dielectric constant of crystal

n = index of refraction

Z_o = characteristic impedance of transmission line

The length of the crystal, L, is directly proportional to the phase retardation, δ . The relationship is as follows

$$L = \frac{\delta \lambda}{\pi n^4 r E_Z} \tag{6-11}$$

Also, it may be observed that the longer the crystal the smaller the field strength required for a given phase retardation.

Figure 6-9 shows a plot of length of crystals required as a function of electric field across the transmission line planes for various modulation indices. Assuming the transmission line to be terminated by a matched load the power (p) required to support these fields is approximately equal to

$$p = \frac{E_z^{\,2} b^2}{Z_o} \tag{6-12}$$

where the attenuation introduced by the electro-optical crystal is being neglected.

This is justifiable if

$$Z_o \omega \epsilon'' \epsilon_o \frac{wL}{b} << 1 \qquad (6\text{-}13)$$

For Z_o = 50 ohms, λ = 0.5 micron, and using a KDP crystal as a phase modulator the power requirements as a function of modulation index are shown in Figure 6-10. The power input is relatively low if the crystal aperture is small. Over the 100 cm length the transmission loss of light through KDP is not negligible. Since the aperture of the modulation is probably smaller than the laser beam width a recollimation of the light would be necessary.

To obtain phase modulation over wide bandwidths with relatively low powers a long modulator is necessary. The size of the modulator described would be approximately 2 by 2 by 100 cm and would weigh about 5 pounds. The same modulator can be used at low or high frequencies. At 100 mc about 30 pounds of extra equipment to produce an output of 10 watts is necessary.

The length of the crystal may be reduced by having the laser beam zig-zag across the microwave beam (Figure 6-11). The angle of incidence is critical in establishing syncronization between the microwave and light beam velocities. To obtain intensity modulation the incident light would have components of polarization in both principal planes of the crystal. In this case the neutral bi-refringence of crystals, like the hydrogen phosphates, limits the optical pathlength severely since the beam is not along the optic axis. In cubic crystals such as ZnS this limitation does not exist.

Fabry-Perot Modulators

One of the mechanisms proposed to enhance the electro-optical properties of Pockels effect crystals is the Fabry-Perot intensity modulator. It promises wide-band modulation at reduced size and power requirements.

Basically it consists of an electro-optical material placed between reflecting surfaces to form a Fabry-Perot resonator (Figure 6-12). If the optical parameters of the electro-optical material are varied even slightly, a considerable change in the transmission of the resonator is produced. This becomes evident if the transmission equation of a lossless Fabry-Perot resonator is considered (Reference 6-7)

$$\frac{I_t}{I_o} = \frac{1}{1 + \frac{4R^2}{(1-R^2)^2} \sin^2\left(\frac{2\pi fLn}{c}\right)} \qquad (6\text{-}14)$$

where I_o and I_t are the respective incident and transmitted light intensities, R^2 is the power reflectivity of the reflectors, f is the optical frequency, L is

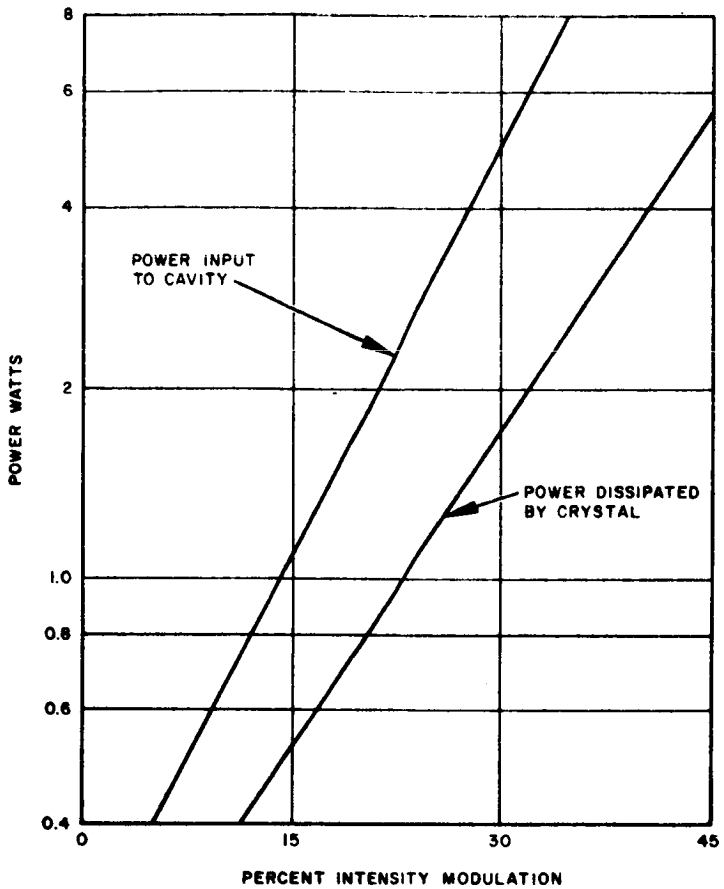


Figure 6-5. Percent Intensity Modulation Versus Power Input and Power Dissipated by Crystal

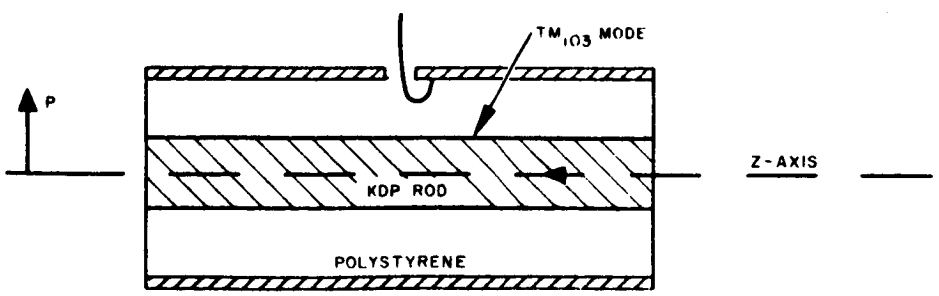


Figure 6-6. Traveling-Wave Modulator

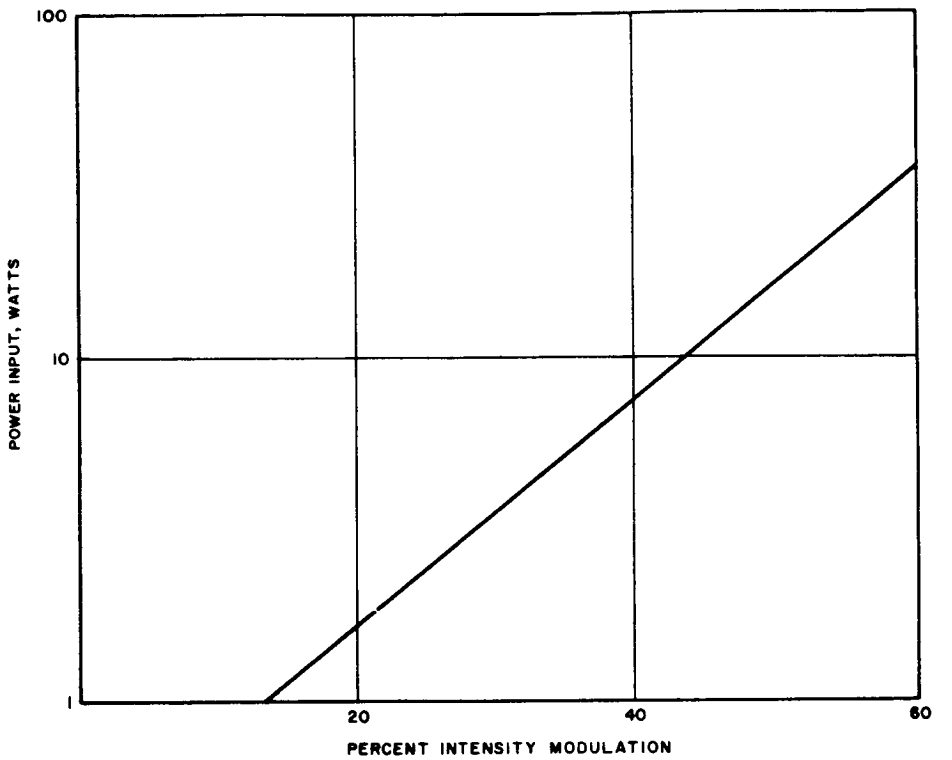


Figure 6-7. Power Input Versus Percent Intensity Modulation for Figure 6-6

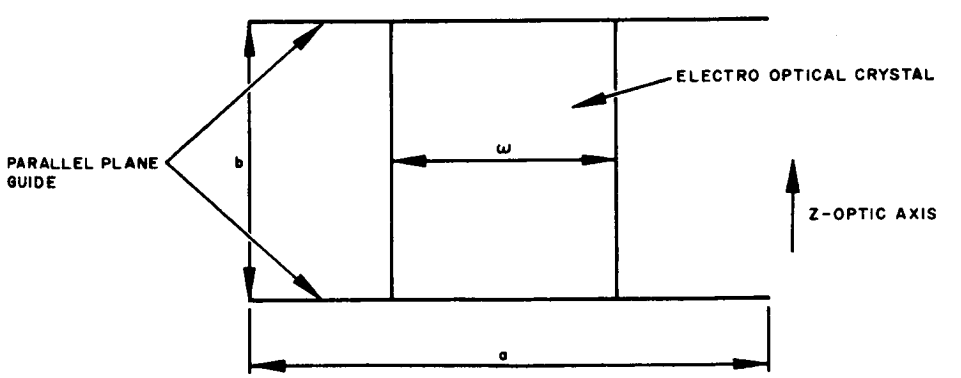


Figure 6-8. Cross section of Traveling-Wave Phase Modulator

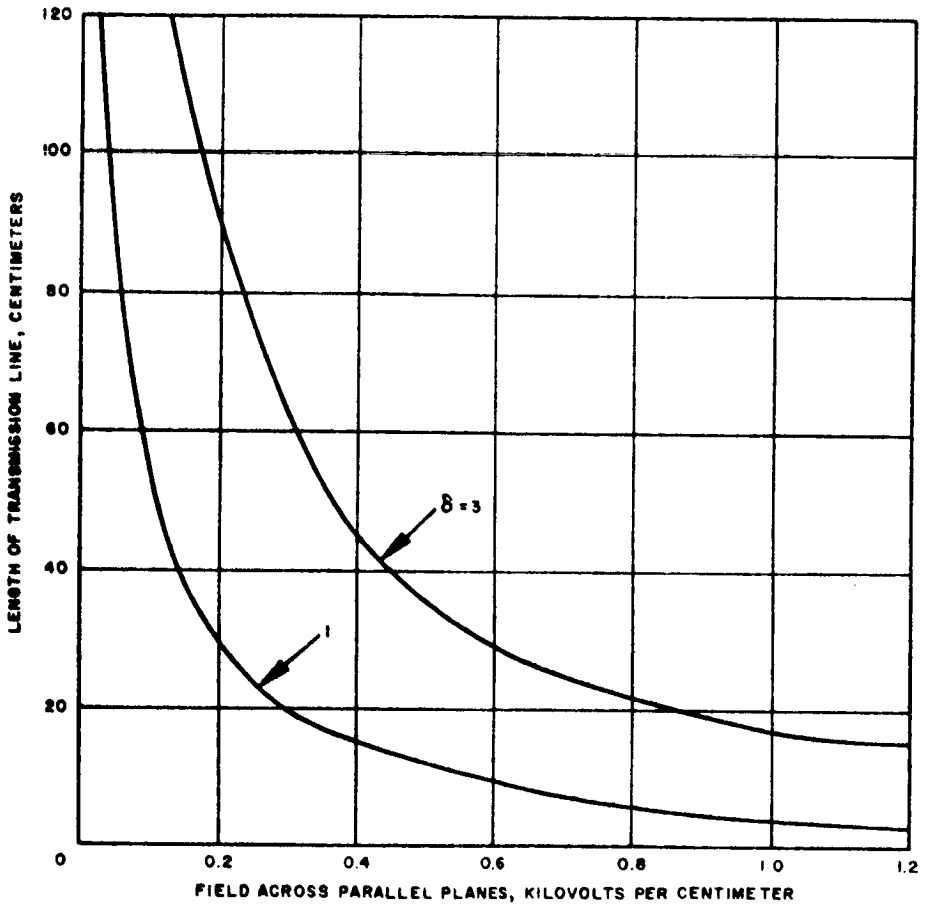


Figure 6-9. Transmission Line Length as Function of Field Across Optic Axis for Modulation Indices

the spacing between the reflectors, n is the index of refraction, and c is the vacuum light velocity. When, through the electro-optic effect, n is varied there will be a corresponding variation in I_t.

The transmission I_t/I_o is plotted as a function of optical frequency f in Figure 6-13. The transmission peaks occur when $2\pi fLn/c = m\pi$ where m is an integer. They are thus spaced at frequency intervals equal to c/2Ln. The peak transmission for the lossless resonator is 1 and the half power frequency interval or passband (Δf) can be obtained from Equation 6-14 by making it equal to 1/2 and solving for f

$$\Delta f \approx \frac{1 - R^2}{R} \quad \frac{c}{2\pi Ln} \tag{6-15}$$

When a varying component $\Delta n \cos \omega t$ is introduced in the index of modulation of the electro-optic material, $n = n_o + \Delta n \cos \omega t$. Then, assuming $2\pi fLn_o/c = m\pi$,

$$\sin^2 \left(\frac{2\pi fLn}{c}\right) = \sin^2 \left(m\pi \frac{\Delta n}{n_o} \cos \omega t\right) \tag{6-16}$$

For practical cases $m\pi\Delta n/n_o < 0.1$, allowing the approximation

$$\sin^2 \left(\frac{2\pi fLn}{c}\right) \approx \left(m\pi \frac{\Delta n}{n_o}\right)^2 \cos^2 \omega t \tag{6-17}$$

Substituting Equation 6-16 into Equation 6-14 yields

$$\frac{I_t}{I_o} \approx \frac{1}{1 + \frac{4R^2}{(1 - R^2)^2} \left(\frac{m\pi\Delta n}{n_o}\right)^2 \cos^2 \omega t} \tag{6-18}$$

Despite changes of $\Delta n/n_o$ in the order of 10^{-6} in the electro optical material with a coefficient of reflectivity $R^2 = 0.9$, the coefficient of $\cos^2 \omega t$ is close to unity leading to large intensity modulation.

Equation 6-17 indicates that the intensity modulation rate is twice the modulation frequency. It can be shown that linear modulation may be obtained by introducing crossed polarizers and a quarter-wave plate bias at

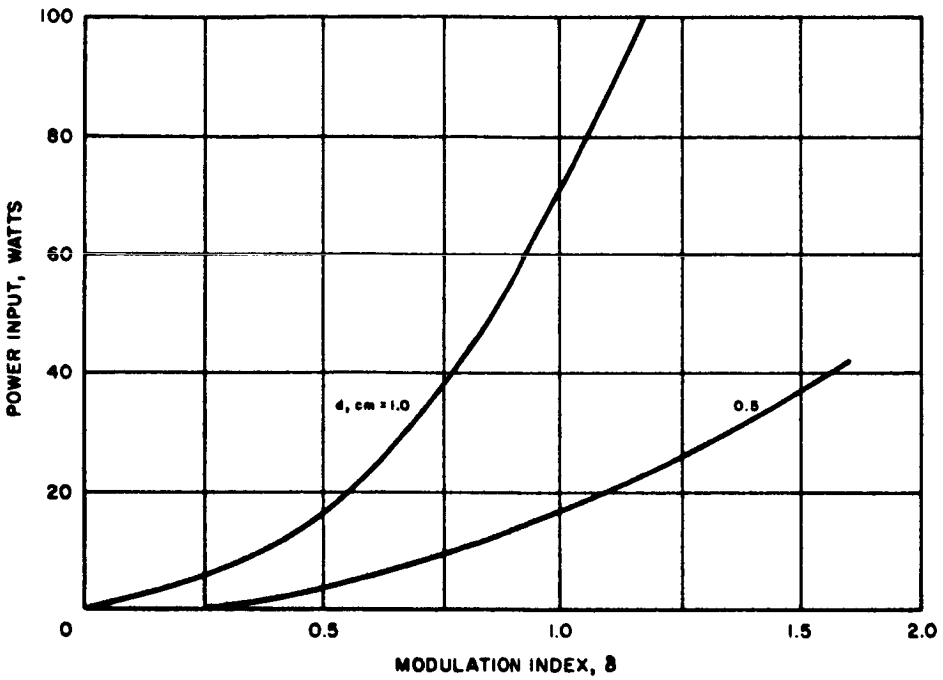


Figure 6-10. Power Input to Phase Modulator as Function of Modulation Index

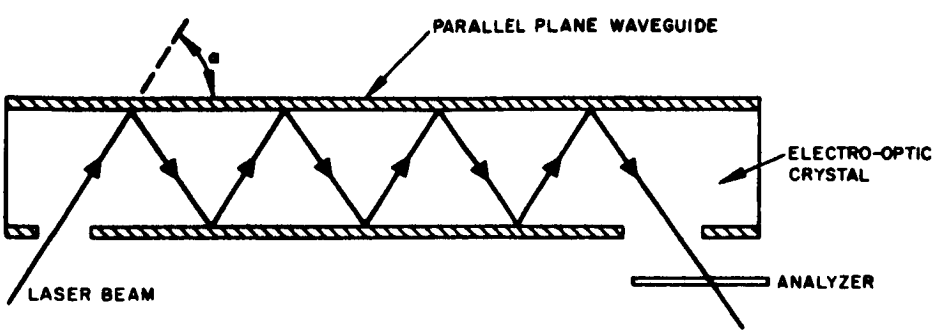


Figure 6-11. Traveling-Wave Light Modulator

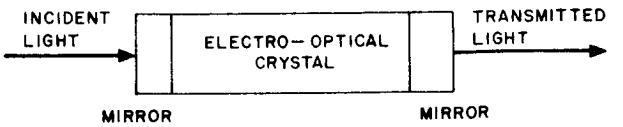


Figure 6-12. Fabry-Perot Resonator-Modulator Schematic

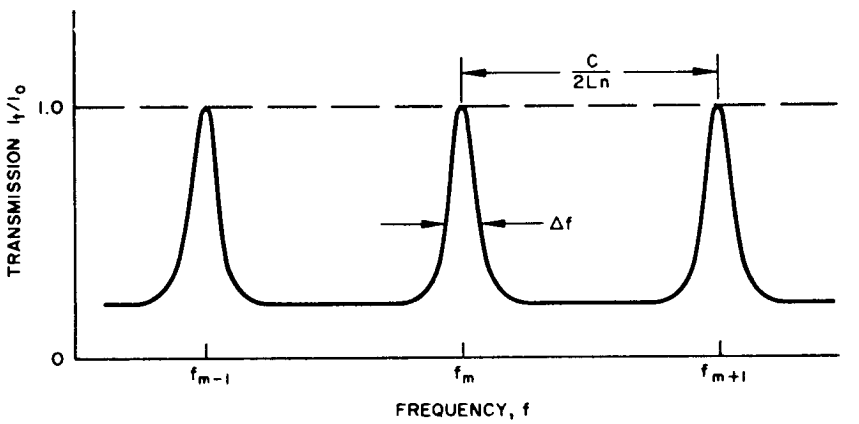


Figure 6-13. Transmission Characteristic of Fabry-Perot Resonator

the ends of the Fabry-Perot resonator. For a lossless resonator, the transmitted intensity then becomes (Reference 6-8)

$$\frac{I_t}{I_o} = \frac{1}{2}\left(\frac{1 + \frac{\pi m}{2}\ \frac{(1+R)^2}{1-R^2}\ \frac{\Delta n}{n}\ \cos \omega t}{1 + \left(\frac{\pi m}{2}\ \frac{(1+R)^2}{1-R^2}\ \frac{\Delta n}{n}\right)^2\ \cos^2 \omega t}\right)^2 \qquad (6\text{-}19)$$

which has a component directly proportional to the change in index of refraction.

The possibility of using a Fabry-Perot structure for intensity modulation of a light beam is thus clear. In practice the Fabry-Perot resonator is not lossless as assumed, but will introduce transmission and absorption losses. If A^2 is the fraction of light absorbed, N^2 the fraction of light transmitted and R^2 the light reflected, the following relationship must hold

$$R^2 + N^2 + A^2 = 1 \qquad (6\text{-}20)$$

Equations 6-14 and 6-19 are then modified by a factor

$$P^2 = \alpha \left(1 - \frac{A^2}{1 - R^2}\right)^2 \qquad (6\text{-}21)$$

resulting in a peak transmission I_t/I_o always less than 1. α is a constant which depends on the flatness of the reflecting surfaces.

A figure of merit of the Fabry-Perot resonator is the contrast difference C defined as

$$C = \left(\frac{I_t}{I_o}\right)_{max} - \left(\frac{I_t}{I_o}\right)_{min} \qquad (6\text{-}22)$$

Substituting Equations 6-21 and 6-14 into Equation 6-22 yields

$$C = P^2\ \frac{4R^2}{(1+R^2)^2} \qquad (6\text{-}23)$$

It becomes quite evident that the reflectivity R^2 plays an important part in the performance of the Fabry-Perot modulator. To obtain high modulation efficiencies the coefficient of $\cos^2 \omega t$ in Equation 6-1 has to be as large as possible. This requires R^2 to be close to unity. On the other hand, the peak transmission and contrast difference must be as large as possible. This excludes a high reflectivity and requires compromise.

Figure 6-14 shows the peak transmission plotted as a function of the coefficient of reflectivity. In order to obtain high transmission, the absorption A^2 of the reflecting surface must be kept low Figure 6-15 demonstrates how critically the transmission depends on A^2/N^2. If, for instance, $A^2/N^2 = 0.05$, the absorption would be less than 0.5 percent when the reflectivity is 0.9, a figure not easily obtainable in practice. The peak transmission is then 0.77. See Table 6-2 for the properties of some of the dielectric coating materials used for Fabry-Perot interferometers (Reference 6-9).

TABLE 6-2. DIELECTRIC COATING MATERIALS

Material	Refractive Index	Loss in Multilayer Stack, percent
Cryolite	1.35	–
ZnS	2.30	0 to 1
Sb_2O_3	2.1	1 to 2
PbF_2	1.8	0 to 1

The other parameter of importance to obtain high transmission is the flatness of the reflecting surface. The flatness is measured as λ/k where λ is the optical wavelength and k is a constant. In practice, the very best plate flatness feasible is $\lambda/100$. However, it is possible that small areas of a flat surface have a k larger than 100. For laser light modulation only small surface areas are required, and therefore a flatness of $\lambda/200$ may be possible. In Figure 6-14 the peak transmission has been plotted for ks equal to 100 and 200. The improvement in transmission is evidenced in flatter surfaces at high values of R^2.

Figure 6-16 shows a plot of the factors which affect the contrast difference. When f(R) and P^2 are multiplied, C is obtained. It would appear that C peaks at approximately $R^2 = 0.7$. For a flatness $k = 200$ and $A^2/N^2 = 0.05$, the peak transmission is 0.9 and the contrast difference is 0.87.

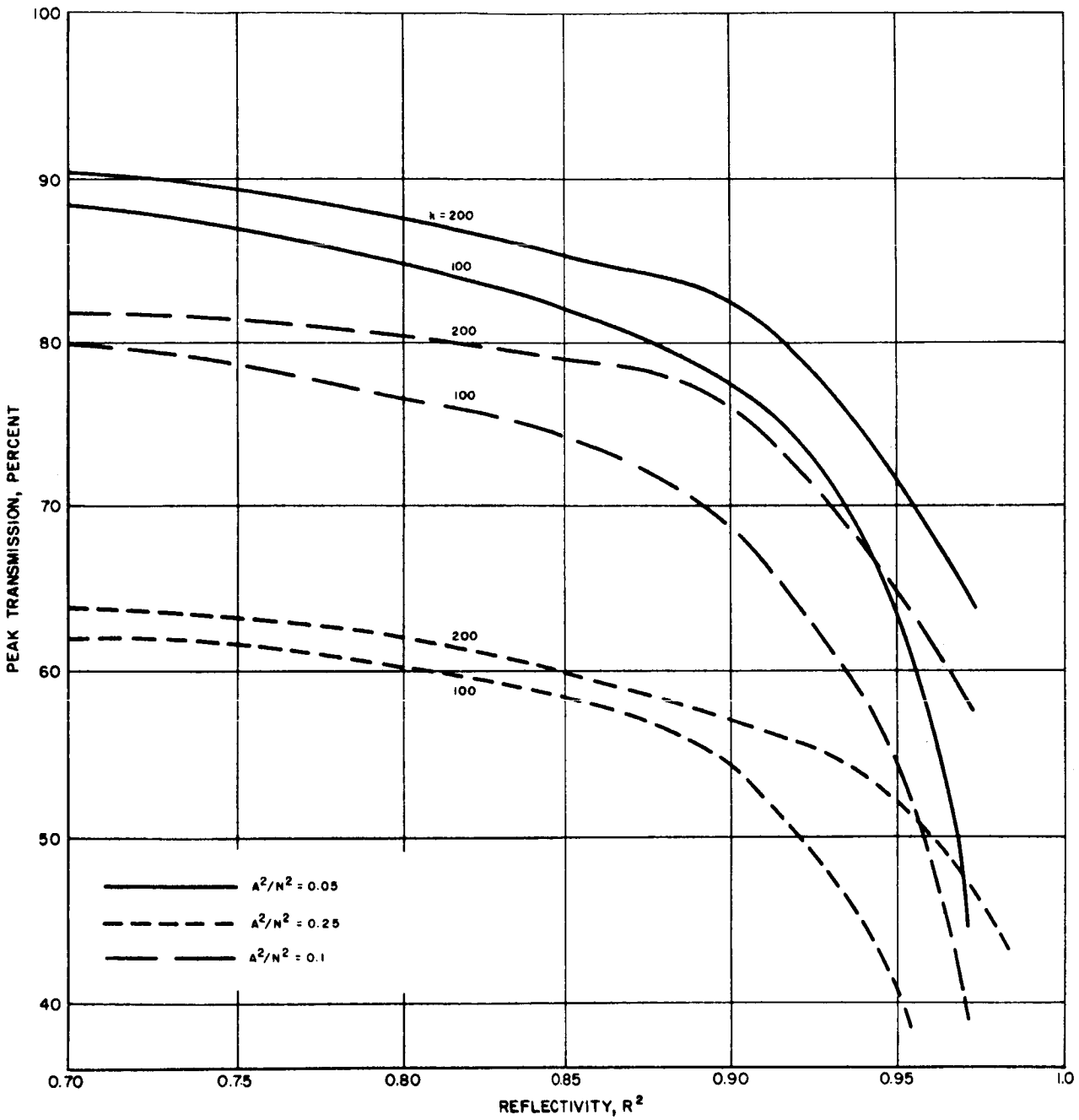


Figure 6-14. Peak Transmission as Function of Reflectivity

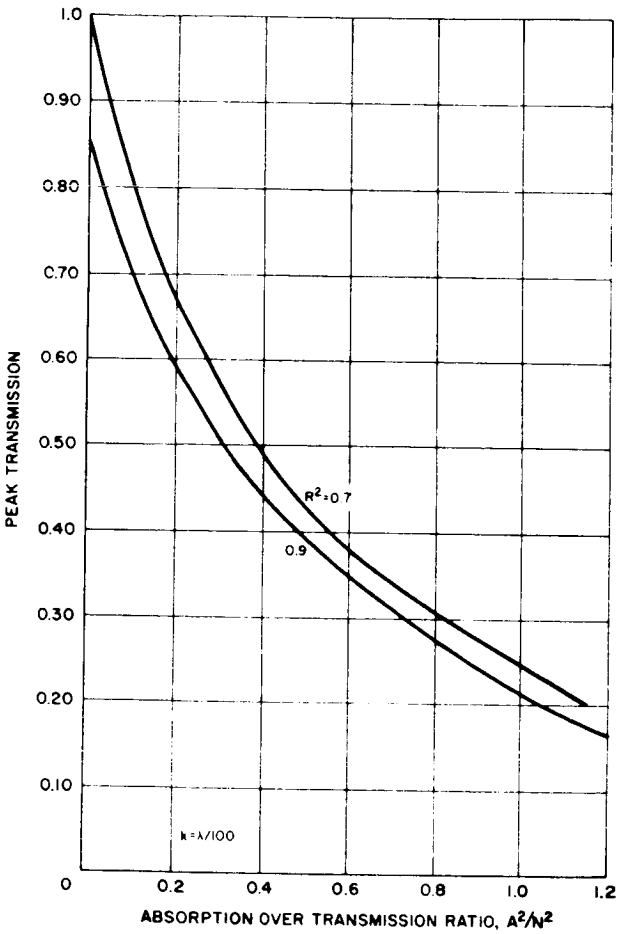


Figure 6-15. Peak Transmission as Function of Absorption-to-Transmission Ratio

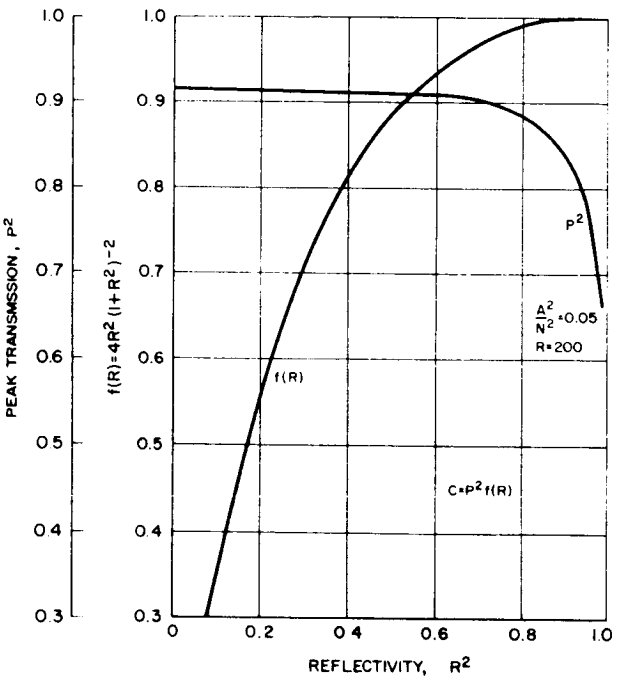


Figure 6-16. Factors Affecting Contrast Difference

In Equation 6-18 the modulation is affected by the reflectivity. The coefficient of $\cos^2 \omega t$ should be as large as possible for maximum modulation efficiency. It is proportional to $g(R^2) = R^2(1-R^2)^{-2}$. Figure 6-17 is a graph of this function. At $R^2 = 0.7$, $g(R^2)$ is only 7.7. At $R^2 = 0.90$ $g(R^2)$ has increased to 90. At this value of R^2 however the peak transmission and contrast difference have decreased to 0.83. It seems reasonable to allow this loss of transmission for the marked improvement in modulation efficiency. Throughout the rest of the discussion on Fabry-Perot modulators a reflectivity of 0.9 will be assumed.

The variation of index of refraction in electro-optical crystals of the XDP type along their privileged axes is given by

$$\Delta n = \pm \frac{1}{2} n_o^3 r E_z \cos \omega t = \pm \frac{1}{2} n_o^3 r \frac{V}{L} \cos \omega t \qquad (6\text{-}24)$$

where $E_z \cos \omega t$ is the electric field applied parallel along the crystal's optic axis and r is the electro-optic constant.

In the following case, polarized light is incident at 45 degrees to the privileged axis of the electro-optical crystal. The crystal is followed by a quarter wave plate and an analyzer crossed with respect to the incident polarization (Figure 6-18). Substituting in Equation 6-19 the transmitted intensity becomes

$$\frac{I_t}{I_o} = \frac{P^2}{2} \left(\frac{1 + B V \cos \omega t}{1 + B^2 V^2 \cos^2 \omega t} \right)^2 \qquad (6\text{-}25)$$

where

$$P^2 = \left(1 - \frac{A^2}{1 - R^2} \right)^2$$

and

$$B = \frac{f \pi (1+R)^2 r n_o^3}{2c (1 - R^2)} \qquad (6\text{-}26)$$

Expanding Equation 6-25 in series form results in

$$\frac{I_t}{I_o} = \frac{P^2}{2} (1 + 2 B V \cos \omega t - B^2 V^2 \cos^2 \omega t - 4 B^3 V^3 \cos^3 \omega t + \ldots) \qquad (6\text{-}27)$$

It is therefore apparent from Equation 6-27 that for minimum distortion B V must be kept much smaller than unity, thereby limiting the magnitude of the modulation index. However, at high modulation frequencies, the distortion is less severe than the equation would indicate since the harmonic content of the transmitted intensity is further attenuated by the frequency selective properties of the Fabry-Perot resonator.

If the modulation is to be in a digital form distortion is of secondary importance. The quarter-wave plate of Figure 6-18 could be omitted in which case the transmission equation becomes (Reference 6-10)

$$\frac{I_t}{I_o} = P^2 \frac{B^2 V^2}{\left[1 + B^2 V^2\right]^2} \tag{6-28}$$

where the parameters are as defined above. The light is 100 percent intensity modulated when B V is varied from 0 to 1, the peak transmission occurring at B V = 1 for which

$$\frac{I_t}{I_o} = \frac{P^2}{4} \tag{6-29}$$

When the analyzer is parallel with respect to the incident polarization (Reference 6-4)

$$\frac{I_t}{I_o} = P^2 \frac{1}{\left[1 + \frac{B^2 V^2}{4}\right]^2} \tag{6-30}$$

Peak transmission occurs at B V = 0 and is equal to P^2. However, minimum transmission is never zero and is limited by the maximum value that B V can practically attain.

In Figure 6-19 the relative transmitted intensity has been plotted as a function of BV for the three types of modulators discussed above. A typical value of B is 10^{-3} (volts)$^{-1}$, indicating a high voltage requirement for large indices of modulation. For small values of BV the largest and most linear variations of intensity occur when crossed polarizers and a quarter-wave plate are used.

Video Intensity Light Modulator

In a Fabry-Perot resonator capable of modulating a light beam up to video frequencies, the type of information to be transmitted is of the form: $v = V \cos \omega t$. The arrangement of Figure 6-18 produces a linear intensity modulation. Disregarding higher order harmonics, from Equation 6-27 the transmission becomes

$$\frac{I_t}{I_o} = \frac{P^2}{2} (1 + 2 BV \cos \omega t) \tag{6-31}$$

The modulation index is defined as

$$x = \frac{1}{P^2} \left[\left(\frac{I_t}{I_o} \right)_{max} - \left(\frac{I_t}{I_o} \right)_{min} \right] = 2 BV \tag{5-32}$$

Substituting Equation 6-26 in Equation 6-32:

$$x = \frac{f \pi r n_o^3 (1 + R)^2}{c (1 - R^2)} V \tag{6-33}$$

Figure 6-20 is a plot of the modulation index, x, versus the applied voltage, V, assuming $P^2 = 0.8$, $f = 5 \times 10^{14}$ cps ($\lambda = 0.6\mu$), $R^2 = 0.9$, and KD*P is used as the electro-optic material ($r = 1.7 \times 10^{-11}$ m/volt, $n_o = 1.468$).

The percent distortion d has also been plotted in Figure 6-20. The distortion is defined here as the ratio of second to first harmonic ($d = 50 B V$). As mentioned above, the distortion curve is somewhat pessimistic since high frequency second harmonics would be further attenuated due to the inherent spectral character of the Fabry-Perot resonator.

The modulation bandwidth is limited by the passband Δf of the resonator given in Equation 6-15. The maximum crystal length is thus fixed by the modulation bandwidth. For KD*P a crystal length L = 4 cm would allow approximately 42 mc bandwidth if $R^2 = 0.9$.

The power required will depend on the electrical circuit used to excite the modulator. Figure 6-21 is such a circuit. R_1 and R_2 are the respective equivalent resistances of the electrodes and the crystal; C is the crystal capacitance, and R_3 is designed to be one order of magnitude larger than the largest impedance of the modulator in order to maintain a constant input impedance as seen from the modulating source.

Assuming a 10 mm^2 cross section and 4 cm length the KD*P ($\epsilon' = 45$) modulator capacitance is

$$C = \epsilon' \epsilon_o \frac{A}{L} = 45 \times 8.85 \times 10^{-14} \quad \frac{0.1}{4} = 0.1 \text{ pf}$$

R_2 is in general approximately 20 times larger than the impedance presented by C. R_1 being very small can be neglected. Thus,

$$R_3 = 0.1 \times \frac{20}{21\ \omega C} = \frac{0.095}{\omega C}$$

The power dissipated is thus

$$p = \frac{V_{rms}^2}{R} = \frac{V_{rms}^2}{R_3} + \frac{V_{rms}^2}{R_2} = 10.55\ \omega C\ V_{rms}^2 \qquad (6\text{-}34)$$

This equation is plotted in Figure 6-22 as a function of the modulation frequency ω for several values of V. It is noted that 1.5 watts are required to produce approximately 57 percent intensity modulation at 42 mc. The relatively high power required is mainly dissipated by R_3 whose only function is to provide a constant impedance looking into the modulator circuitry. If R_3 could be avoided, the power requirements would be drastically reduced.

The temperature of the crystal changes when power is applied to it. Assuming a uniformly heated rod which is long compared to its diameter, the temperature rise ΔT can be written

$$\Delta T = \frac{1}{4}\ \omega \epsilon_o \epsilon''\ \frac{A}{L^2}\ \frac{V^2}{K} \qquad (6\text{-}35)$$

where K is the thermal heat conductivity and ϵ'' is the reactive complex permittivity. For KD*P, $K \simeq 10^{-2}$ watts/cu^{o}K and $\epsilon'' \simeq 0.01\ \epsilon'$, yielding a temperature rise of 0.04^{o}K at 100 mc and 65-percent intensity modulation for the crystal dimensions discussed above (A = 10 mm^2, L = 4 cm). Changes

in temperature will tend to change the index of refraction and the length of the cavity, a serious effect at high modulation frequencies. Up to video frequencies it is estimated that temperature changes up to $1^{o}K$ can be tolerated.

Microwave Intensity Light Modulation

In Equation 6-14 and Figure 6-12 the transmission of the Fabry-Perot resonator peaks at frequency intervals f_i equal $c/2\ Ln_o$. Now the modulation frequency is chosen to be f_i. Under such circumstances the sidebands of the intensity modulated light, f, would occur at $f_o \pm f_i$ and as a result would be transmitted by the resonator. It can be shown that the sideband amplitude is proportional to the modulation index, x, for small x. The power in the first sideband is proportional to x^2 which is proportional to the microwave modulating power, and the bandwidth capability is approximately Δf. Thus it is possible to obtain microwave intensity modulation at a frequency f_i and bandwidth Δf.

The microwave field E_z applied along the crystal's z-axis is now no longer constant as in the case of the video modulator. It takes the form of a cosinusoidal standing wave as shown in Figure 6-23. The standing wave may be broken down into a forward and backward traveling wave, each moving at the velocity of light in the crystal and having an amplitude $E_z/2 \cos \omega_i t$. Light traveling in the same direction as the microwave field sees a constant field amplitude and thus suffers a phase retardation. On the other hand, light traveling in opposite directions to the microwave field sees a varying field whose average effect is to cancel out phase retardations. Since the change of index of refraction is proportional to the field, Equations 6-24 and 6-26 are modified by a factor of 1/2.

Modulation at microwave frequency does not cause an increase in power requirements as compared to the video modulator previously discussed. The reason for this is that the modulation power of the video modulator is mainly dissipated in R_3, the resistance required to maintain a constant load impedance. At microwave frequencies cavities may be tuned at the modulation frequency. The minimum bandwidth of the microwave cavity will be limited by the dielectric Q of the electro-optic crystal. The power requirements are then mostly contributed by the crystal's power dissipation.

The power dissipated by the crystal is

$$p = \frac{1}{4}\ \omega_i\ \epsilon_o\ \epsilon''\ \frac{A}{L}\ (E_z\ L)^2 \qquad (6\text{-}36)$$

where ω_i is the microwave modulation frequency fixed by the length L of the Fabry-Perot resonator, ϵ_o is the free space dielectric constant, ϵ'' is the reactive permittivity constant of the crystal, A is its cross-sectional area, and E_z is the applied peak microwave electric field. The factor 1/4 arises from a time and space average of the microwave standing wave pattern.

The relationship between the applied field and the modulation index is now

$$x = 1.07\ 10^{-3} E_z \tag{6-37}$$

Assuming parameters similar to those of the video modulator, the required modulating frequency is found to be

$$\omega_i = 2\pi \frac{c}{2 L n_o} = 2\pi \times 0.256 \times 10^{10} \text{ radians/sec} \tag{6-38}$$

when L = 4 cm. At this frequency, assuming a crystal dielectric Q of 100, the minimum microwave cavity bandwidth is 25.6 mc. To obtain 50 mc modulation the cavity would have to be loaded to a Q of 50. For this to hold, it is assumed the complex permittivity has increased to a value $\epsilon'' = \epsilon'/50$, which for KD*P is approximately 0.90. Substituting these values into Equation 6-36 yields

$$p = 0.435\ x^2 \text{ watts} \tag{6-39}$$

which is plotted in Figure 6-24. Only 150 milliwatts are required to obtain 60 percent intensity modulation at 2.5 gc.

Polarization Modulators

In polarization of the light transmitted by the Fabry-Perot modulator, the incident light may be linearly polarized at 45 degrees to the privileged x and y axes of the electro-optic crystal. Under these circumstances the optical vectors transmitted can be shown to be

$$A_x = \frac{A_o P}{\sqrt{2}\ \sqrt{1+(BV)^2}} \sin\left(2\pi f t - \tan^{-1} BV\right)$$

$$A_y = \frac{A_o P}{\sqrt{2}\ \sqrt{1+(BV)^2}} \sin\left(2\pi f t + \tan^{-1} BV\right) \tag{6-40}$$

where $A_o \sin 2\pi f t$ is the incident polarized optical vector, V is the applied voltage, P is the transmission and as previously defined

$$B = \frac{\pi f n^3 r}{2c} \quad \frac{(1+R)^2}{(1-R^2)} \tag{6-41}$$

The transmitted optical vector thus becomes elliptically polarized due to the phase difference between A_x and A_y. To restore linear polarization, A_x and A_y should be out of phase by $m\pi$ (m is an integer), that is $2 \tan^{-1} BV = m\pi$. This would evidently require an infinite value of BV. Circular polarization, however, is possible, when $2 \tan^{-1} BV = m\pi/2$. Then $BV = 1/\sqrt{2}$ for m = 1, and the required voltage can be easily calculated. When $f = 5 \times 10^{14}$ cps, n = 1.5, $r = 1.7 \times 10^{-11}$ m/v, and $R^2 = 0.9$, by substitution into Equation 6-41 $B = 5.4 \times 10^{-3}$ volts $^{-1}$ and V = 130 volts. The circular polarization is right- or left-handed depending on the polarity of V. Unfortunately, the transmitted intensity does not remain constant but is also dependent on BV. From Equation 6-14 it can be shown that

$$\frac{I_t}{I_o} = \frac{P^2}{1+(BV)^2} = 0.53$$

assuming circular polarization and $P^2 = 0.8$.

Digital communications are most suitable for polarization modulation. For instance, 1 and 0 bits could be characterized by right- and left-hand circular polarizations of the light beam.

For an idealized digital polarization modulator the voltage, current, and power waveforms in the electrical circuit have been drawn in Figure 6-25. The bit rate S is equal to T^{-1}, the inverse of the waveform time period. Assuming a 4 cm long, 0.1 cm^2 cross section, KD*P Fabry-Perot modulator, the capacitance is 0.1 pfd, and the voltage necessary for circular polarization is 130 volts. The average power requirements are then

$$P_{av} = CV^2 S = 1.75 \times 10^{-9} S \tag{6-42}$$

and if T = 0.1 T the peak power is

$$P_m = 20 CV^2 S = 3.5 \times 10^{-8} S \tag{6-43}$$

Equations 6-42 and 6-43 are plotted in Figure 6-26.

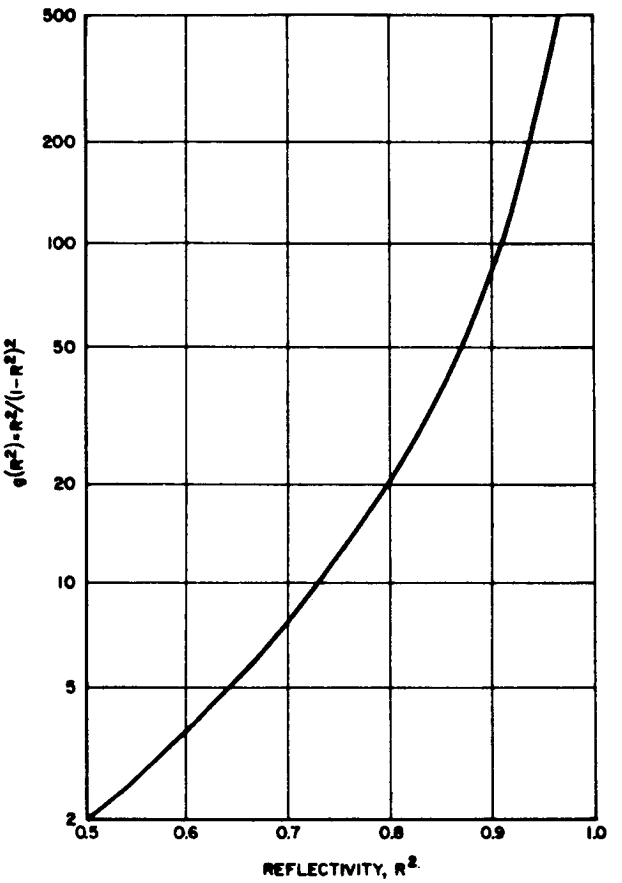


Figure 6-17. Coefficient of Modulation Factor

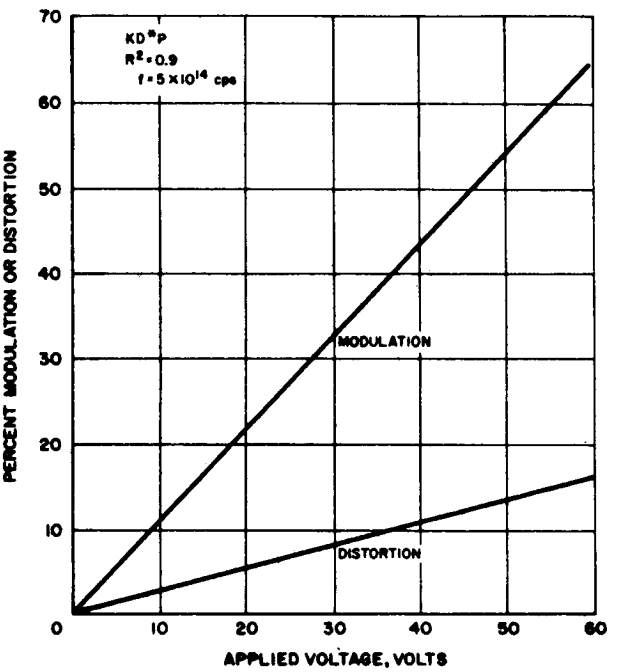


Figure 6-18. Percent Modulation and Distortion Versus Peak Applied Voltage for Linear Intensity Modulation

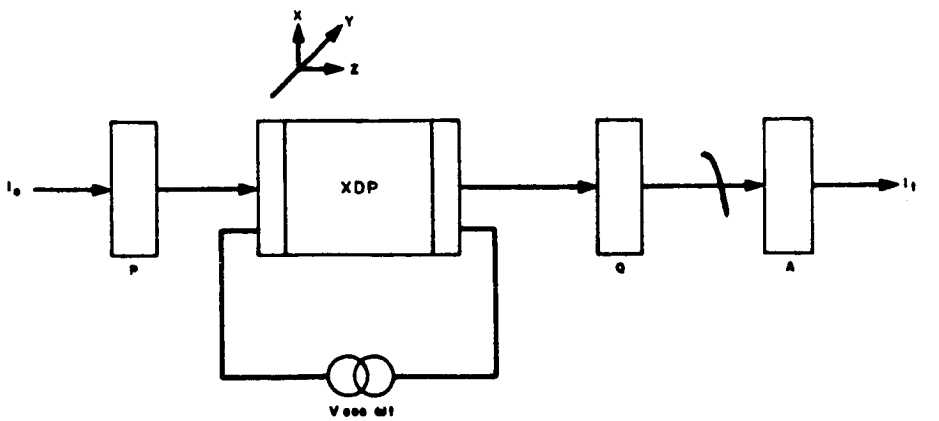


Figure 6-19. Fabry-Perot Linear Intensity Modulator Schematic

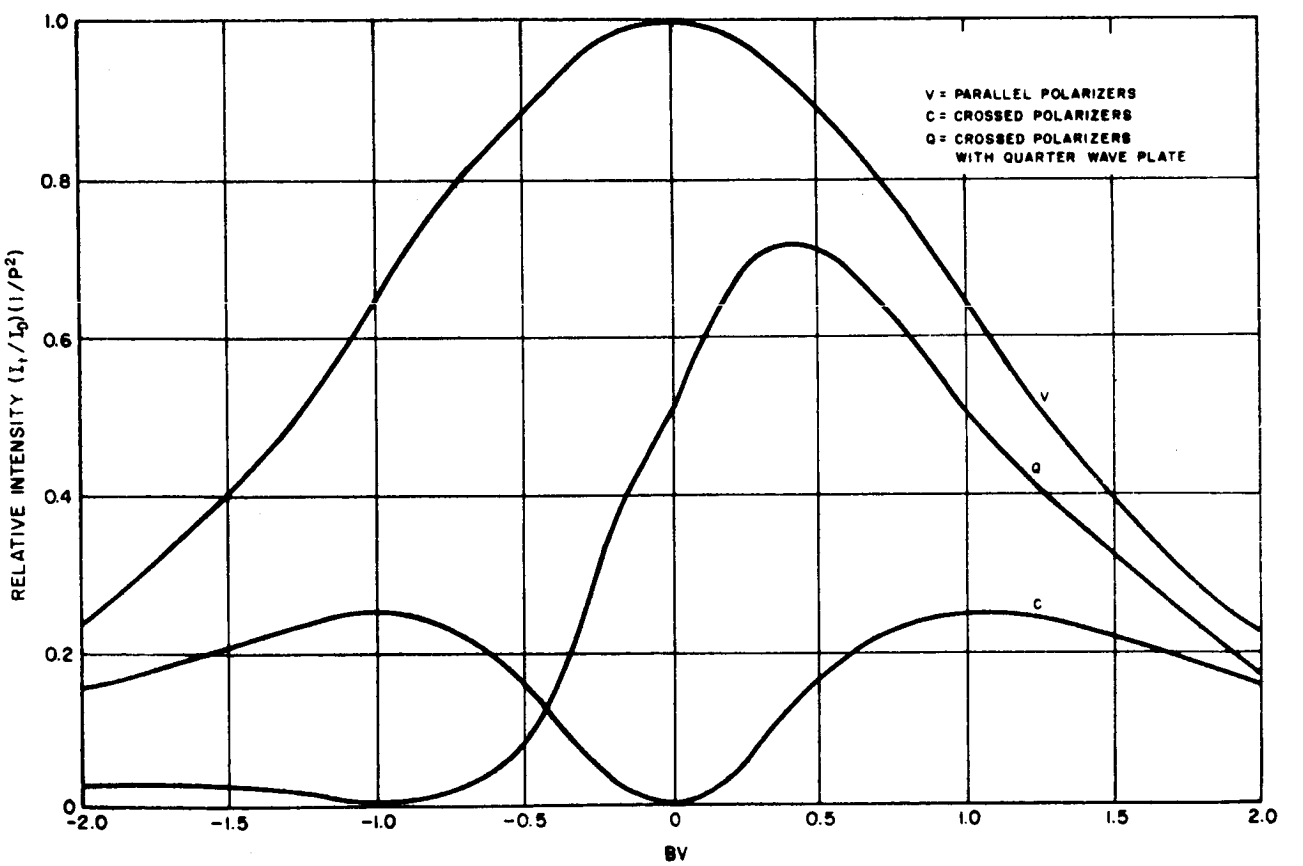


Figure 6-20. Relative Intensity Versus BV for Various Types of Fabry-Perot Modulators

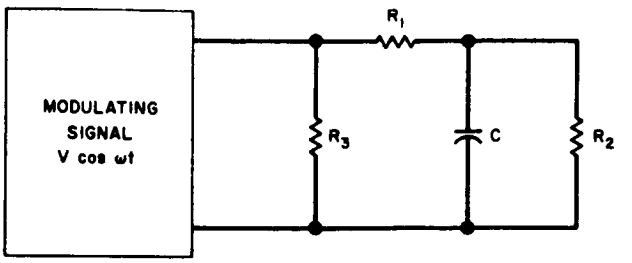


Figure 6-21. Equivalent Circuit of Fabry-Perot Resonator

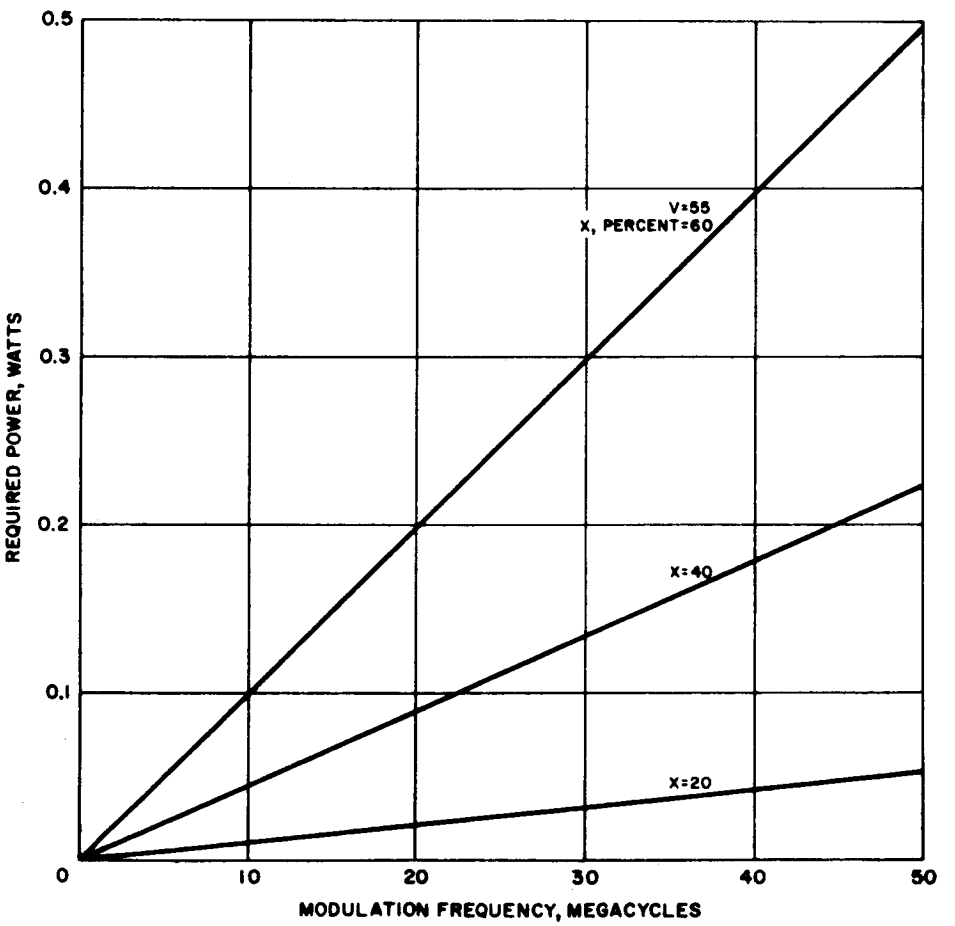


Figure 6-22. Power Required Versus Frequency for Different Percentage Values of Intensity Modulation

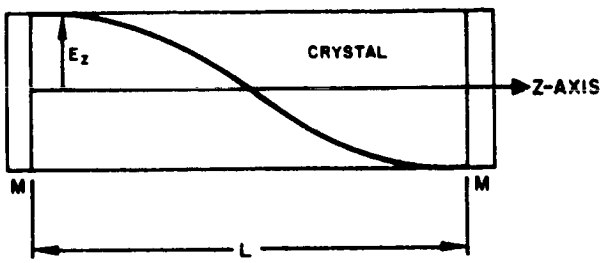


Figure 6-23. Standing Wave Field Applied to Fabry-Perot Microwave Modulator

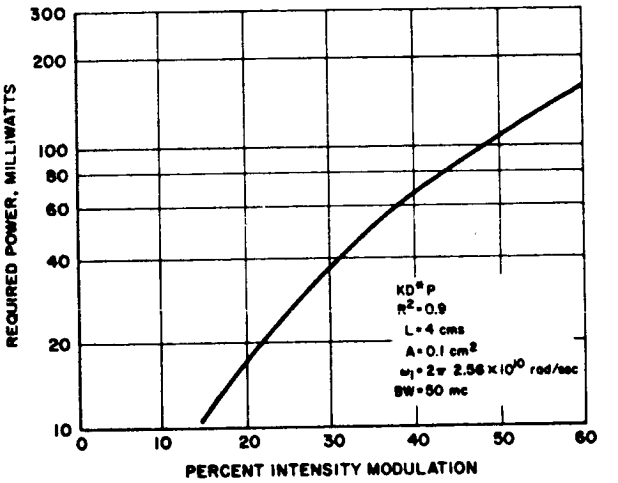


Figure 6-24. Required Power Versus Modulation of Fabry-Perot Microwave Modulator

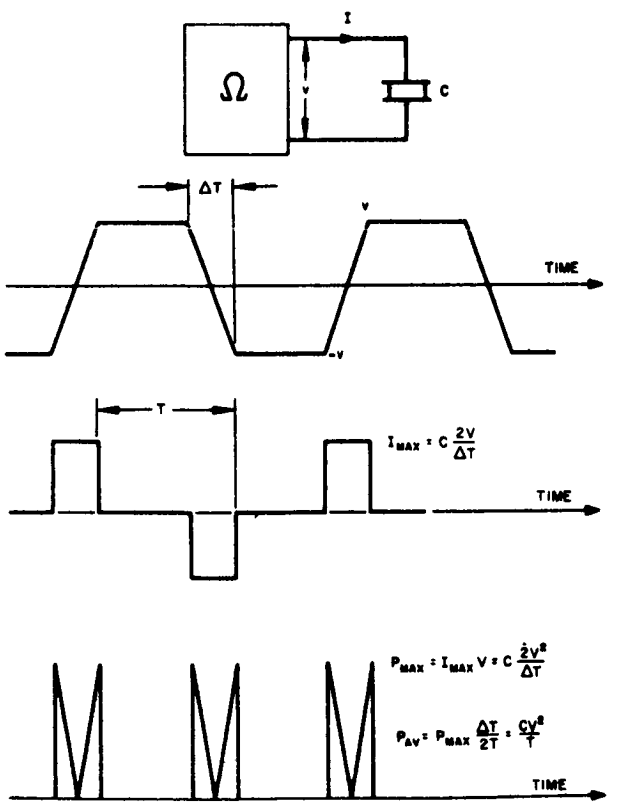


Figure 6-25. Voltage, Current, and Power Waveforms in Idealized Digital Polarization Modulator

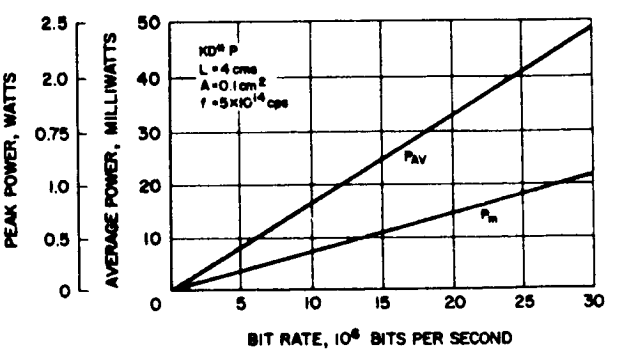


Figure 6-26. Average and Peak Power Versus Bit Rate for Circular Polarization Modulation in Fabry-Perot Resonator

Fabry-Perot Phase Modulator

The phase retardation obtained in an electro-optic crystal when subjected to an electric field is a consequence of its change of index of refraction, which in turn governs the velocity of light propagation through the medium. Equation 6-40 showed that in a Fabry-Perot structure a beam polarized along the privileged x-axis of the crystal will be transmitted as

$$A_{tx} = \frac{P A_{ox}}{\sqrt{1 + B^2 V^2 \cos^2 \omega t}} \sin (2\pi f t - \delta) \qquad (6\text{-}44)$$

where $\delta = \tan^{-1} (BV \cos \omega t)$ is the phase retardation, $A_{ox} \sin 2\pi f t$ is the incident polarized optical vector, and $V \cos \omega t$ is the impressed voltage along the z axis.

For values of $BV < 0.4$ the approximation $\delta \simeq BV \cos \omega t$ can be made with a 5 percent error and linear phase modulation is possible. Note that since a phase retardation is necessary along only one of the privileged axes, the direction of light propagation need not be along the z axis as in the case of intensity and polarization modulation. It could be along the y axis, for instance, as long as the plane of polarization was parallel to the x axis and the electric field of course was maintained parallel to the z axis. The electric field along the z axis in this case becomes

$$E_z = \frac{V}{d}$$

where d is the crystal length in the z direction. The electric field accounts for the change in index of refraction. Assuming video modulation, the parameter B now becomes

$$B = \frac{\pi f n^3 r L (1+R)^2}{2 c d (1 - R^2)} \qquad (6\text{-}45)$$

which is an improvement of L/d over Equation 6-26.

As in the case of polarization modulation the transmitted intensity is reduced by

$$\frac{I_t}{I_o} = \frac{P^2}{1 + B^2 V^2} = 0.86 P^2 \qquad (6\text{-}46)$$

for BV = 0.4. The voltage required is then only 7.4 volts when L = 4 cms, d = 0.4 cms, and the variation of intensity is 14 percent. If Equation 6-44 is expanded into its harmonics by making use of Bessel functions and the approximation

$$\frac{1}{\sqrt{1 + B^2V^2 \cos^2 \omega t}} = 1 - \frac{B^2V^2}{2} \cos^2 \omega t \qquad (6\text{-}47)$$

the following is obtained:

$$A_{tx} = \left(1 - \frac{(BV)^2}{4}\right) J_o (BV) \sin 2\pi f t + J_1 (BV) \left[\pm\left(1 - \frac{(BV)^2}{4}\right) \pm \frac{1}{2} \right]$$

$$\sin (2\pi f \pm \omega)t + \frac{(BV)^2}{4} \quad \frac{J_o (BV)}{2} \sin (2\pi f \pm 2\omega)t \qquad (6\text{-}48)$$

where harmonics less than 0.01 have been neglected. Even for BV = 0.4 the first sideband is less than 11 percent of the carrier and the second sideband is less than 2 percent. The modulation is quite weak, even when the bandpass of the phase modulator is designed to accept the second sideband of the modulation. For modulation bandwidths up to 25 mc a 4-cm long Fabry-Perot resonator may be used. The light is propagated along the y axis and is polarized along the x axis. The voltage along the z axis is impressed across the thickness of the crystal (0.4 cm). Figure 6-27 shows the modulator. The power requirements at video frequencies are computed making electrical circuit assumptions similar to video intensity modulation. From Equation 6-34 the power dissipated is

$$p = 10.55 \ \omega C \ V_{rms}^2 \qquad (6\text{-}49)$$

which is plotted in Figure 6-28 for different indices of modulation (i.e., values of BV). The power requirements are seen to be only a few milliwatts, an insignificant amount when compared with other types of modulation.

By introducing the Fabry-Perot resonator into a microwave structure the modulation frequency may be increased to that of the frequency interval between the transmission peaks of the resonator as previously discussed. The modulation sidebands are made to coincide with the transmission peaks. The modulation bandwidth however, is still limited by the resonator's

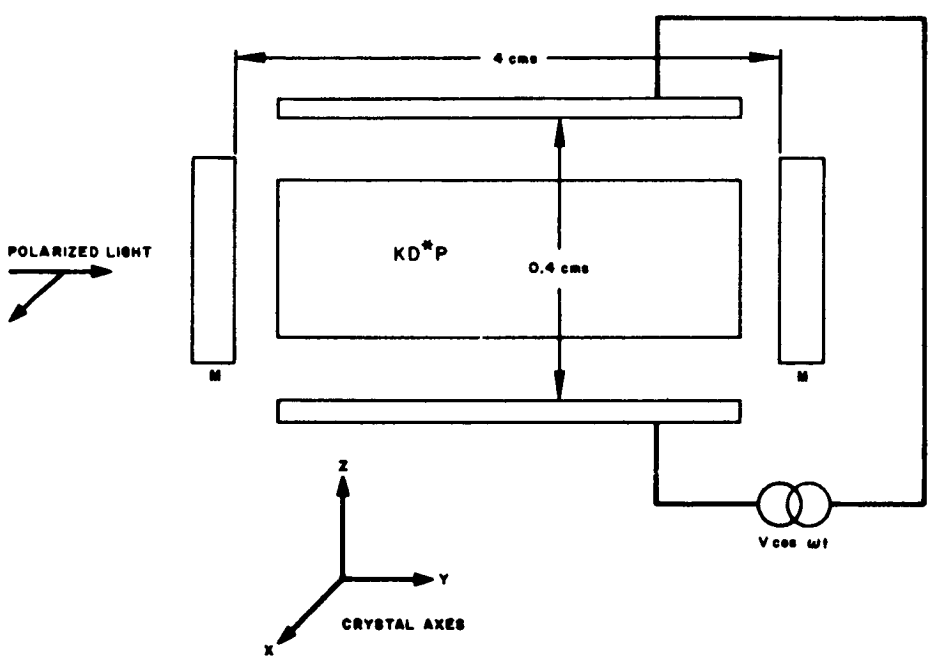


Figure 6-27. Fabry-Perot Phase Modulator

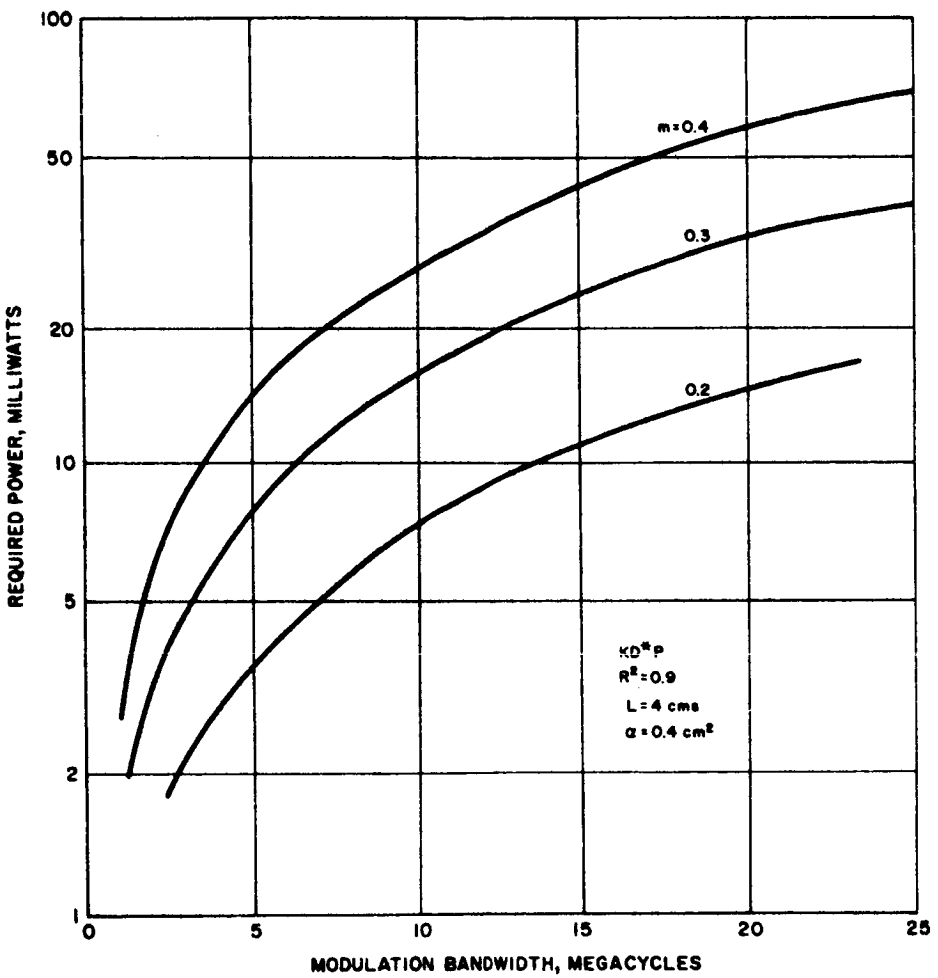


Figure 6-28. Modulation Bandwidth (mc)

passband. The power requirements may be calculated using Equation 6-36 which only takes into account crystal losses. Assuming L = 4, the modulation frequency is 2.5 gc and

$$p = \frac{1}{4}\,\omega \epsilon_o\, \epsilon''\, L\, \frac{m^2}{2B^2} = 144\, m^2 \text{ milliwatts}$$

which for modulation indices $m \leq 0.4$ is always less than 25 milliwatts.

If digital information is to be transmitted, the phase modulation need not be linear and larger values of BV can be tolerated. However, phase shifts equal to $\pi/2$ are not possible because of its dependence on $\tan^{-1} BV$. A phase shift equal to $\pi/3$, for instance would require 32 volts. The transmitted intensity would be 20 percent.

Comparison of Electro-Optical Modulators

The Fabry-Perot modulator imposes two important constraints on the laser light which are not present in other types of modulators. First, it must oscillate in only one axial mode or, if not that, the axial modes must be separated by at least Δf. The linewidth of the laser emission must of necessity be orders of magnitude smaller than Δf. In addition, the beam must be well collimated and perpendicular to the reflecting surfaces. A diverging beam would transmit the typical Fabry-Perot fringes. To ensure that most of the light is transmitted in the central fringe, the beam spread should be smaller than 10 minutes of arc, a requirement readily met by most lasers.

The Fabry-Perot modulator requires considerable care and skill in its construction. Its transmission is sensitive to flatness and absorption. The laser beam must be kept as narrow as possible so that a surface flatness of k = 200 can be found. Parallelism between the plates is also essential.

The performance of two types of electro-optic intensity light modulators can be compared in two cases:

1) The electro-optical crystal is within a Fabry-Perot resonator.

2) The electro-optical crystal is not within a Fabry-Perot resonator.

In both cases light is assumed polarized at 45 degrees to the privileged axis of the crystal and followed by a quarter-wave plate and analyzer crossed with respect to the incident polarization.

In the case of the Fabry-Perot modulator, the light transmission up to video frequencies is

$$\frac{I_t}{I_o} = \frac{P^2}{2}\left(1 + \frac{f \pi\, r\, n_o^3}{c}\,\frac{(1+R)^2}{(1-R^2)}\, V \cos \omega t\right) \qquad (6\text{-}50)$$

whereas in case 2 the transmission has been shown to be

$$\frac{I_t}{I_o} = \frac{T^2}{2}\left(1 + \frac{2f\pi r n_o^3}{c} V \cos \omega t\right) \qquad (6\text{-}51)$$

The factor of improvement between cases 1 and 2 can be measured by the ratio of the coefficients of cos ωt in Equations 6-50 and 6-51

$$\rho = \frac{P^2}{T^2}\frac{(1+R)^2}{2(1-R^2)} \qquad (6\text{-}52)$$

T^2 is caused by reflection losses at the surface of the crystal due to improper matching and for small crystal lengths is usually 0.9 for practical cases. P^2 is given by Figure 6-14 and depends on the reflection coefficient R^2. ρ is significantly dependent on the reflection coefficient due to the $(1 - R^2)$ factor in the denominator of Equation 6-51. For $R^2 = 0.9$, we obtain $\rho \simeq 16$, which indicates that for the same modulation index, the Fabry-Perot resonator consumes $(T^4/P^4)\ \rho^2 \simeq 361$ less power. The advantage is somewhat offset by the greater distortion introduced when the modulation index is large.

As discussed previously, the modulation bandwidth of the Fabry-Perot resonator is limited by its length. Long lengths, on the other hand, are desirable because the capacitance of the crystal is reduced and thus the power requirements decrease. The length of the crystal does not limit the bandwidth in case 2. For equal power requirements and bandwidth, the length in case 2 would have to be ρ^2 times longer than the Fabry-Perot resonator. However, the difficulties associated with fabricating a long rod of electro-optical material are great, and over long lengths the losses in light transmission through the crystal are considerable. The transmission as a function of length is roughly

$$\frac{I_t}{I_o} = \exp\left(-\frac{L}{35}\right) \qquad (6\text{-}53)$$

when L is in centimeters.

Since the power requirements in case 2 are large, the bandwidth and modulation index will ultimately be limited by how much power the electrodes and the crystal can dissipate. It will be assumed that 0.5 watt is the maximum power a 4 cm x 0.1 cm^2 crystal can dissipate the maximum bandwidth and modulation index for a KD*P crystal are then related by $x^2\ \omega = 2\pi\ 28 \times 10^6$. The bandwidth is thereby limited to 28 mc for 100 percent modulation.

To compare intensity light modulators at microwave frequencies, the microwave Fabry-Perot and traveling-wave intensity modulators previously discussed will be considered.

Using the linear intensity modulation configuration of Figure 6-19 for the traveling-wave modulator, the transmitted intensity is

$$\frac{I_t}{I_o} = \frac{T^2}{2}\left(1 + \frac{2f}{c}\ \pi r\, n_o^{\,3}\, E_z L_t \cos \omega t\right) \tag{6-54}$$

where a microwave mode has been assumed such that

$$\int_o^{L_t} E_z\, dl = E_z L_t \tag{6-55}$$

E_z being the electric field applied along the crystal's z-axis, L_t the length of the traveling wave structure, and T^2 the light transmittance through the crystal.

In the case of the Fabry-Perot modulator

$$\frac{I_t}{I_o} = \frac{P^2}{2}\left(1 + \frac{f\pi\, r n_o^{\,3}}{c}\ \frac{(1+R)^2}{2(1-R^2)}\ E_z L_f \cos \omega t\right) \tag{6-56}$$

where again an assumption similar to Equation 6-55 has been made. Comparing Equation 6-56 and 6-54 the following figure of merit is obtained:

$$\rho = \frac{P^2}{T^2}\ \frac{(1+R)^2\ L_f}{4(1-R^2)\ L_t} \tag{6-57}$$

The light transmission for KDP is expressed in Equation 6-53. Thus for $R^2 = 0.9$

$$P^2 = 0.8 \exp(-L_f/35)$$

$$T^2 = \exp(-L_t/35)$$

Substituting into Equation 6-57

$$\rho = 7.6 \frac{L_f}{L_t} \exp\left[(L_t - L_f)/35\right] \tag{6-58}$$

Figure 6-29 shows a plot of ρ as a function of the length of the traveling-wave modulator, L_t, for lengths of the Fabry-Perot modulator L_f equal to 1 and 4 cm. For $L_f = 4$ cm, ρ is always larger than 2 regardless of the length of the traveling-wave modulator. The power requirements for both modulators become equal when $\rho^2 (T^2/P^2)^2 = 1$, which corresponds in this case to $L_t = 38$ cm, assuming similar impedance matches in the microwave structures. The Fabry-Perot modulator however is limited in modulation frequency, modulation bandwidth, and index of modulation whereas the traveling-wave modulator is limited only in how well the microwave and light velocities can be matched. For $L_f = 4$ cm the modulation bandwidth of the Fabry-Perot modulator is 50 mc. If L_f is reduced to 1 cm the modulation bandwidth of the Fabry-Perot resonator is increased to 330 mc. The power requirement then becomes twice as large as that required for a 19-cm traveling-wave modulator at the same modulation index. The transmission for a 19-cm traveling-wave modulator is $T^2 \simeq 0.57$, whereas that of the 1-cm Fabry-Perot modulator is $P^2 \simeq 0.80$.

In the case of polarization modulation, a linearly polarized light vector, $A_o \sin 2\pi f t$, will be incident upon a modulator and inclined at a 45-degree angle to the x-y axes of the electro-optic crystal. When the Fabry-Perot resonator is removed, the transmitted optical vectors are

$$A_x = T \frac{A_o}{\sqrt{2}} \sin(2\pi f t - \delta)$$

$$A_y = T \frac{A_o}{\sqrt{2}} \sin(2\pi f t + \delta) \tag{6-59}$$

where the phase retardation, $\delta = \pi f n^3 r V/c$. The incident linearly polarized beam becomes elliptically polarized. When the total phase difference between the vectors is $2\delta = \pi/2$, circular polarization results. When $2\delta = \pi$, the emergent beam is linearly polarized and crossed with respect to the incident polarization. The voltage required for circular polarization is hence

$$V = \frac{c}{4 f n^3 r} \tag{6-60}$$

When a Fabry-Perot resonator is used, the voltage required is (from Equation 6-41)

$$V = \frac{1 - R^2}{(1+R)^2} \quad \frac{2c}{\pi f n^3 r} \quad \frac{1}{\sqrt{2}} \tag{6-61}$$

Comparing Equations 6-60 and 6-61, it can be noted that the Fabry-Perot voltage requirements are smaller by a factor of

$$\rho = \frac{0.54\,(1+R)^2}{(1 - R^2)} = 20 \tag{6-62}$$

for $R^2 = 0.9$.

The considerable power advantage of the Fabry-Perot modulator should be weighed against its transmission loss. The transmission of the Fabry-Perot modulator is approximately 0.50 for circular polarization, whereas without the resonator the transmission is 0.9 or better. The multi-element cascade method can be used to reduce the modulator driving voltage. As described above, the electro-optic crystals are mounted in series with their x-y privileged axes alternately rotated 90 degrees. The polarity of the voltage impressed across each crystal is also alternately reversed. In the case of digital polarization modulation, the voltage necessary to produce circular polarization is thereby reduced by a factor, 1/m, where m is the number of crystals used. Assuming the usual parameters, 2600 volts will produce the quarter-wave retardation necessary for circular polarization when one crystal is used without the Fabry-Perot resonator. If the number of crystals is increased to m = 26, the voltage is reduced to 100 volts. The total capacitance of the device is

$$C = 26 \quad \frac{\epsilon_o \epsilon' A}{L} = 34.5 \text{ pfd}$$

where $A = 0.1 \text{ cm}^2$, $L = 0.3$ cm, and $\epsilon' = 45$.

The average and peak powers required can now be obtained by substituting into Equations 6-42 and 6-43

$$P_{av} = CV^2S = 3.45 \times 10^{-7} S \tag{6-63}$$

$$P_m = 20\,CV^2S = 6.9 \times 10^{-6} S \tag{6-64}$$

These equations are plotted in Figure 6-30. The power requirements of the Fabry-Perot modulator are still smaller by a factor of 50 when Equations 6-42 and 6-44 are compared.

There is no reason why a cascaded modulator of the type described above could not be fitted into a Fabry-Perot resonator. For an m-element modulator, the voltage required for circular polarization would be thereby reduced by a factor m. The capacitance of the device however, is increased by m^2.

$$C = m \frac{\epsilon_o \epsilon' A}{L/m}$$

so that the power requirements remain equal (see Equation 6-42). Nevertheless, the very low voltage necessary for circular polarization when m is large and the relatively low power required make this an attractive method of modulation when a 50 percent loss in transmission can be tolerated.

When light is polarized along the x or y axis of an electro-optic crystal the phase retardation is

$$\delta = \frac{\eta n^3 f r L V_z}{c d} \qquad (6\text{-}65)$$

where V_z is the applied voltage along the z axis, L is the crystal length in the direction of light propagation and d is the crystal length across which V_z is applied. For phase modulation the electric field and the direction of light propagation need not coincide. Comparing Equation 6-65 to the phase retardation in a Fabry-Perot modulator (Equation 6-44),

$$\rho = \frac{(1+R)^2 L_f}{2(1-R^2) L_t} = 19 \frac{L_f}{L_t} \qquad (6\text{-}66)$$

where $R^2 = 0.9$. The Fabry-Perot phase modulator thus has significant voltage and power advantage over the same phase modulator with the Fabry-Perot resonator removed. This advantage however is limited by the unavoidable intensity modulation, the low modulation index for linear modulation, and the relatively narrow bandwidth of the Fabry-Perot resonator. When the Fabry-Perot resonator is removed these disadvantages do not exist and the modulation powers can be made comparable by making the length of the electro-optical crystal longer.

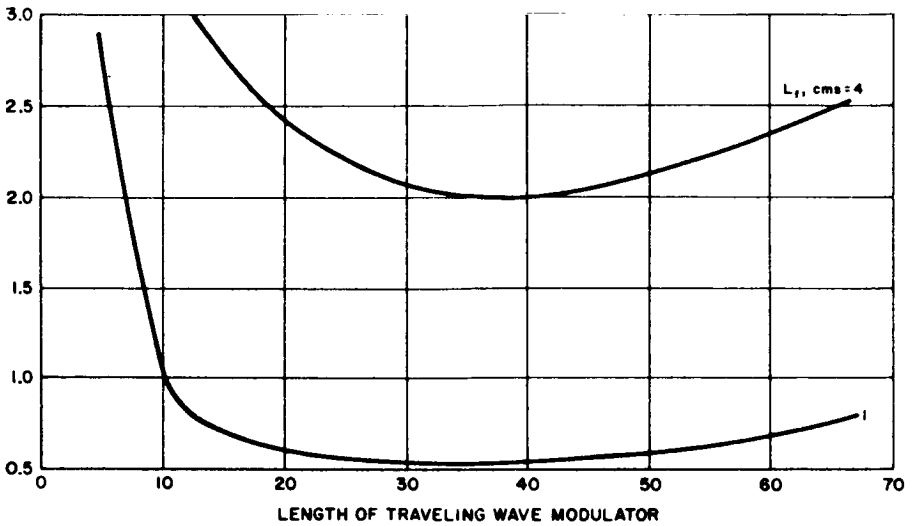


Figure 6-29. Comparison of Microwave Traveling-Wave Modulator and Fabry-Perot Modulator

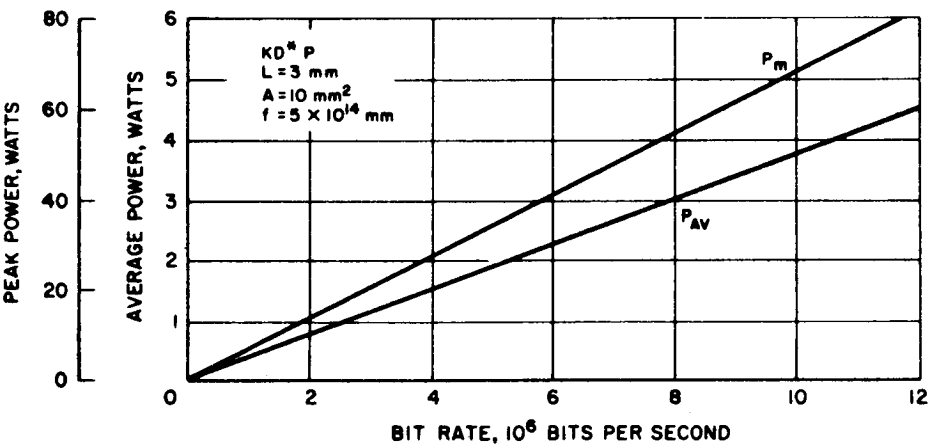


Figure 6-30. Average and Peak Power Versus Bit Rate for Circular Polarization Modulation 26-Element Cascaded Modulator

Conclusions

The most important and perhaps single advantage of the Fabry-Perot modulator is its low voltage and power requirements. Regardless of the type of modulation the voltage necessary is at most 10 volts and the power requirements never exceed 0.5 watts. This would indicate that the information to be transmitted could be applied directly to the modulator from the data processing system, thereby minimizing the size and weight of the equipment. However, the merits of the Fabry-Perot modulator must be weighed against its significant shortcomings. First, the light transmitted is attenuated. The minimum attenuation hoped for seems to be 20 percent, which requires very flat surfaces and very low absorption coating materials. Second, the quality of the modulation is poor in that considerable distortion is introduced when the communication system requires linear modulation. Third, the laser beam must be well collimated, operate in a stable single mode, and have a line width orders of magnitude smaller than the Fabry-Perot passband. These disadvantages do not exist or are not as pronounced in the other mechanism discussed. However, the power requirements are generally two orders of magnitude greater. This means that additional equipment in the form of voltage and power amplifier is required to convert the modulation information from the data processing system to a sufficiently high voltage and power level to drive the modulator.

Table 6-3 shows a comparison between modulators which differ by whether a Fabry-Perot resonator is employed or not. The different parameters are computed on the basis of using KD*P as the electro-optical crystal and light at a 0.6 micron wavelength. The different modulators are not necessarily designed for optimum performance but simply to serve as guide and a means for comparison. Under total equipment size, weight, and power requirements a very rough estimate has been made. These figures do not include prime power equipment.

TABLE 6-3. COMPARISON BETWEEN MODULATORS

Type of Modulation	Video Modulation Frequency, < 50 mc		Microwave Modulation Frequency, > 1 gc	
	Fabry-Perot	Not Fabry-Perot	Fabry-Perot	Traveling Wave
Phase				
Transmission loss, percent	20	10	20	43
Maximum modulation index	0.4	> 1	0.4	> 1
Distortion, percent	5	0	5	0
Maximum modulation bandwidth, megacycles	21	45	21	300
Size, centimeters	4 by 0.4	4 by 0.4	4 by 0.4	20 by 0.4
Modulation frequency, gigacycles			2.56	2.56
Voltage required, volts	7.4	142[+]		
Power required, milliwatts	60[+]	0.0225[+]	25[++]	90[++]
Total equipment size, inches	3 by 2 by 5	12 by 12 by 12	6 by 6 by 12	6 by 6 by 12
Total power required, watts	0.1	100	0.2	0.2
Total weight, pounds	< 0.5	< 20	< 3	< 3
Remarks	The intensity is also modulated.	Bandwidth and modulation index are limited by power dissipated in crystal.		Modulation frequency and bandwidth depend on microwave structure and how well velocities are matched.
Intensity				
Transmission loss, percent	20	10	20	43
Maximum modulation, percent	60	100	60	100

TABLE 6-3 (continued)

Type of Modulation	Video Modulation Frequency, < 50 mc		Microwave Modulation Frequency, > 1 gc	
	Fabry-Perot	Not Fabry-Perot	Fabry-Perot	Traveling Wave
Intensity (continued)				
Distortion, percent	< 15	< 1	< 15	< 1
Modulation frequency, gigacycles			2.56	2.56
Maximum modulation bandwidth, megacycles	42	78	42	300
Length, centimeters	4	4	4	20
Voltage required, volts rms	2.6*	47*		
Power required, watts	0.42	140	0.16**	0.58**
Total equipment size, inches	3 by 2 by 5	12 by 12 by 12	6 by 6 by 12	6 by 6 by 12
Total power required, watts	1	2	2	10
Total weight, pounds	< 1	20	< 3	< 3
Remarks	Modulation bandwidth limited by length of modulator	Modulation bandwidth limited by power dissipated in crystal.	Modulation frequency and bandwidth depend on length.	Modulation frequency and bandwidth depend on how well light and microwave velocities are matched.
Digital Polarization				
Transmission loss, percent	50	10		
Polarization	Right- or left-hand circular	Right- or left-hand circular		

TABLE 6-3 (continued)

Type of Modulation	Video Modulation Frequency, < 50 mc		Microwave Modulation Frequency, >1 gc	
	Fabry-Perot	Not Fabry-Perot	Fabry-Perot	Traveling Wave
Digital Polarization (continued)				
Maximum bit rate, megabits/second	42	15		
Length, centimeters	4	8		
Voltage required, volts	5 †	100 †		
Average power required, milliwatts	25 ††	5500		
Peak power required, watts	0.6 ††	105 ††		
Total equipment size	3 by 2 by 5 inches	1 foot^3		
Total power required, watts	0.2	10		
Total weight, pounds	< 1	< 10		
Remarks	Bit rate limited by length of modulator	Bit rate limited by power dissipated in crystal		

+0.4 modulation index, 21 mc bandwidth.
++0.4 modulation index.
*15 cascaded elements.
**60 percent modulation, 2.56 gc.
†26 cascaded elements.
††15 megabits/second.
†††60 percent modulation, 42 mc.

REFERENCES

6-1. R. W. McQuaid, Applied Optics, 2, March 1963, p. 320.

6-2. J. E. Gensic, Applied Phys. Letters, 2, 15 May 1963, p. 185.

6-3. T. Moreno, "Microwave Transmission Design Data," Dover Publications, Inc., New York, 1948, pp. 227-240.

6-4. W. W. Rigrod and I. P. Kaminow, Proc. IEEE, January 1963, p. 137.

6-5. A. G. Siegman, Applied Physics Letters, November 1962.

6-6. C. J. Peters, "Traveling Wave Modulator," Proc. IEEE, January 1963, p. 47.

6-7. F. A. Jenkins and H. E. White, Fundamentals of Optics, New York, McGraw-Hill, 1950, p. 258.

6-8. E. I. Gordon and T. D. Rigden, "The Fabry-Perot Electro-Optic Modulator," BSTJ, January 1963, p. 174.

6-9. P. W. Baumeister, "Handbook of Optical Design, No. 141," (U. S. Government Printing Office, Washington 25, D. C., 1963) and S. P. Davis, Applied Optics, 2, July 1963, p. 730.

6-10. E. I. Gordon and J. D. Rigden, "The Fabry-Perot Electro-Optic Modulator," BSTJ, January 1963, p. 174.

APPENDIX A6. LASER INJECTION DIODE SIGNAL PROCESSING

The limitations of utilizing a laser injection diode for pulse position modulation transmission in high information rate systems are due both to physical limitations of the device itself and also to the limitations in the laser diode signal processing equipment. These problems are discussed on the following pages.

PPM TRANSMITTING SYSTEM

A conceptually simple method for accomplishing the conversion and operating a laser injection for PPM transmission is shown in Figure A6-1. The operation is as follows: An information signal is gated by a clock pulse so that it charges a small capacitor which discharges through a resistor to produce an exponentially decreasing voltage, the initial value of which is the information signal voltage sample. The time for this voltage to decay to some fixed value is proportional to the logarithm of the input signal. The voltage on the capacitor is used to hold an avalanche transistor in the cut-off condition. Just after the input sample is applied to the small capacitor, the B+ voltage is applied via another gate to the energy storage capacitor of the laser and to the avalanche transistor. The decay of the stored information signal voltage eventually will allow the avalanche transistor to conduct, and the laser will flash. The time delay of the laser pulse is thus proportional to the logarithm of the information signal voltage.

PPM RECEIVING SYSTEM

The signal processing in the receiver would be somewhat similar. If an exponentially rising voltage (i.e., a positive exponent) could be generated, e.g., if the exponential buildup of an oscillator could be utilized, direct inverse processes could be employed. However, it appears best to work from the complementary time interval. At the time of arrival of the signal pulse a capacitor voltage decay can be initiated; then the voltage present on the capacitor at the end of the sample period (less any necessary reset period) represents the transmitted signal. A clock pulse gates out the information signal by sampling this exponentially decaying capacitor voltage at the correct instant. This technique is illustrated in Figure A6-2.

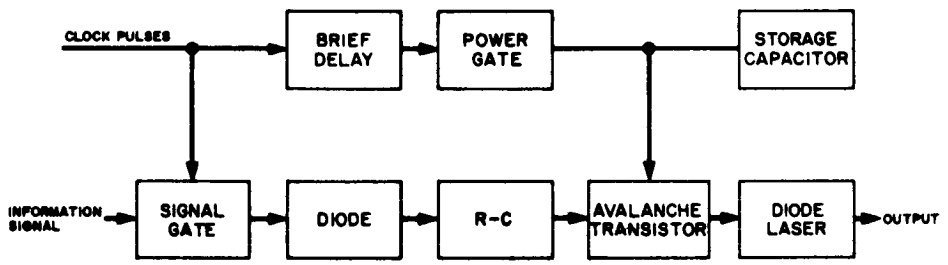


Figure A6-1. Method of Generating Pulse Position Modulated Laser Beam

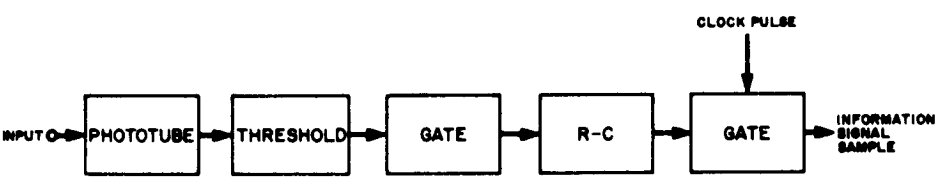


Figure A6-2. Method of Recovering Information Signal from Pulse Position Modulated Beam

SIGNAL PROCESSING LIMITATIONS

Some of the difficulties involved in signal processing a diode laser for PPM transmission are discussed below. The analysis presented is for the basic system outlined previously. As a design example the following system requirements have been chosen for a reduced data rate, real-time TV communication link. They correspond to the PPM system discussed in the Fourth Interim Report of the Deep Space Optical Communications Study.

1) Transmission of 5×10^7 nautical miles.

2) Sampling rate of video signal, 2 mc.

3) Pulse width of laser output, 5 nanoseconds.

4) Average power output, 12 watts.

5) Duty cycle, 1 percent.

The basic approach to this problem has been to restrict the system to all requirements except the power output and then to maximize this power output.

Laser Injection Diode

A few basic assumptions and their limitations must be made about the diode laser before proceeding further. It will be assumed that the diode is Gallium Arsenide, having an energy gap of ~ 1.5 electron volts with 100 percent quantum efficiency. An average power output of 12 watts calls for a 1.2-kilowatt pulse, which in turn requires 800 amperes through the diode during the 5-nanosecond pulse.

Extrapolating from the present figures on existing diode lasers, the following limitations for the above assumptions are found:

1) To produce diode lasers capable of 800 amperes, fabricating techniques must be improved to make the p-n junction area larger by a factor of 30.

2) Since present diode lasers have a bulk resistance of approximately 1 ohm, the widened junction area would reduce this to 0.03 ohm. However, even with such a small resistance, the average power dissipation is about 200 watts; compared to the 12-watt power output, there is a 6 percent efficiency.

3) Most important is the requirement demanded of a cooling system to conduct away 200 watts of power from the bulk resistance of the diode laser.

Pulse Operation of Avalance Switch (See Figure A6-3)

The avalanche switch shown in Figure A6-1 utilizes the following sequence of operations:

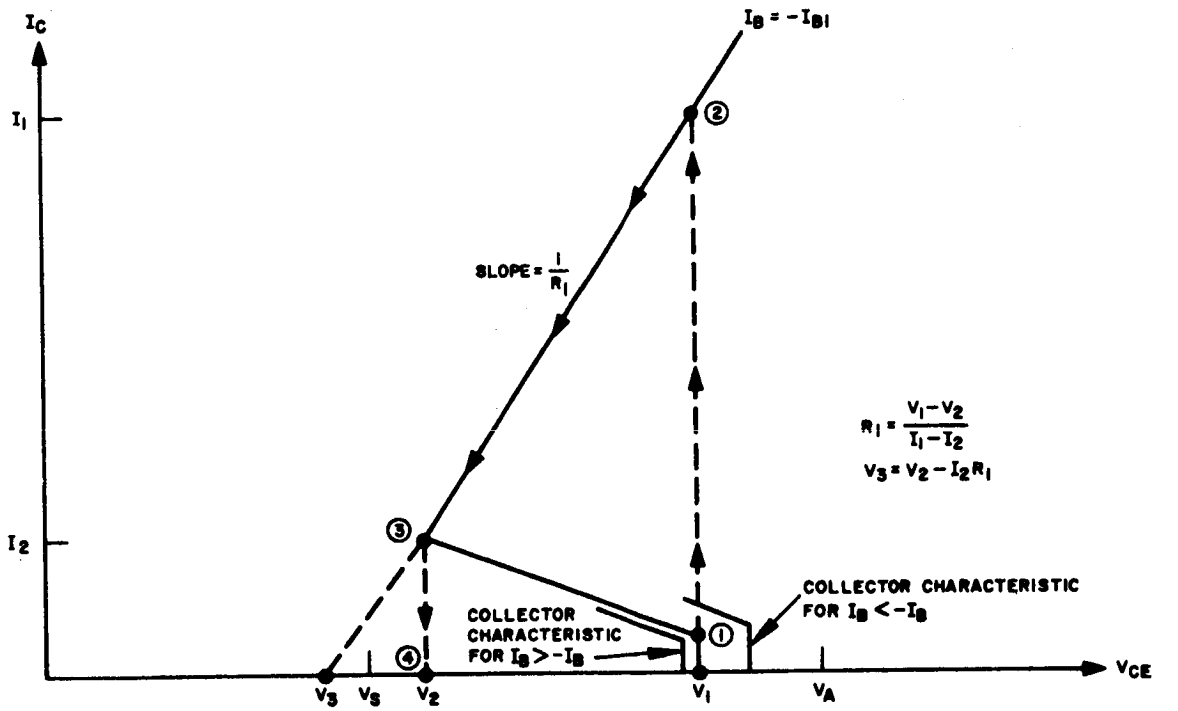
1) Initially the avalanche transistor operates at ($V_{CE} = V_2$, $I_C = 0$). The video sample is gated into C_v so that $I_B < -I_{B1}$, biasing the transistor into cutoff. The supply voltage is then gated across the storage capacitor, C_s, charging it up to V_1; the locus of operation is now ($V_{CE} = V_1$, $I_C = 0$). As the voltage across C_v decays, so does the base current I_B until it eventually reaches a threshold current $-I_{B1}$; the locus of operation now switches up to position 1.

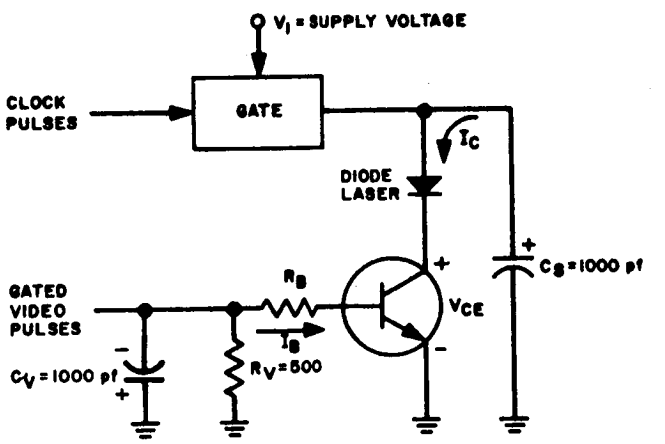
2) As the base current continues to decay below $-I_{B1}$, the knee of the collector characteristic is moved below V_1, and the locus of operation must switch up to position 2. Physically speaking, before this threshold is reached, the base current supplies a reverse saturation current for both the collector and the emitter diodes. As the base current reaches $-I_{B1}$, it can no longer supply a reverse saturation current for the emitter diode; thus, the emitter diode becomes forward biased, injecting minority carriers into the base. These carriers diffuse into the collector junction where avalanche multiplication begins to take place. Some of the additional current flows through R_B, increasing the forward bias across the emitter junction. This in turn increases the carriers injected into the collector for further multiplication until position 2 is reached. The switching time is limited only by external circuitry.

3) At position 2 there is effectively a voltage supply V_3 in series with a resistor R_1 and a capacitor C_s, which is charged up to V_1; thus, there is a resulting R-C decay to position 3, the time of this transition being given by t_1 in Figure A6-3.

4) At position 3 the collector current I_2 forces the voltage of the capacitor to decay below V_2, forcing the locus of operation off the characteristic and down to position 4. Again the time of transition is limited only by external circuitry.

In general, the operation of the switch involves collector voltages of 20 volts or more, which means that the transistor must dissipate 20 to 50 times the power output of the diode laser. This reduces the efficiency of the system by at least another factor of 20. If a single avalanche switch is to be utilized in the above operation described, and if a laser output power of 12 watts is to be achieved, transistor technology will have to be greatly improved with respect to higher power capabilities and possibly lower avalanche voltages.





FOR TRANSITION FROM ② TO ③:

$$\begin{cases} V_{CE}(t) = V_3 + (V_1 - V_3)e^{-t/R_1C_3} \\ V_{CE}(t_1) = V_2 = V_3 + (V_1 - V_3)e^{-t_1/R_1C_3} \\ t_1 = (R_1C_3)\,\mathcal{L}n\left|\frac{V_1 - V_3}{I_2R_1}\right| = 5\times10^{-9}\ \text{SEC} \end{cases}$$

Figure A6-3. Pulse Operation of Avalanche Switch

Control Gates

With present-day technology, it is extremely difficult to design a power gate that will switch high currents in times of nanoseconds. The best gate that can be achieved is a transistor, switching between saturation and cutoff, supplying 500 milliamperes in a switching time of 150 nanoseconds. If 100 nanoseconds is included for charging the capacitor, there is a total "deadtime" of 250 nanoseconds due to the power gate alone (compared to the sampling time of 500 nanoseconds).

To avoid this deadtime difficulty, two such systems may be energized alternately, thus allowing a 500-nanosecond charging time for the storage capacitor in each system as shown in Figure A6-4. With such a configuration, however, both diode lasers must be focused in the same direction. This will require additional optical equipment.

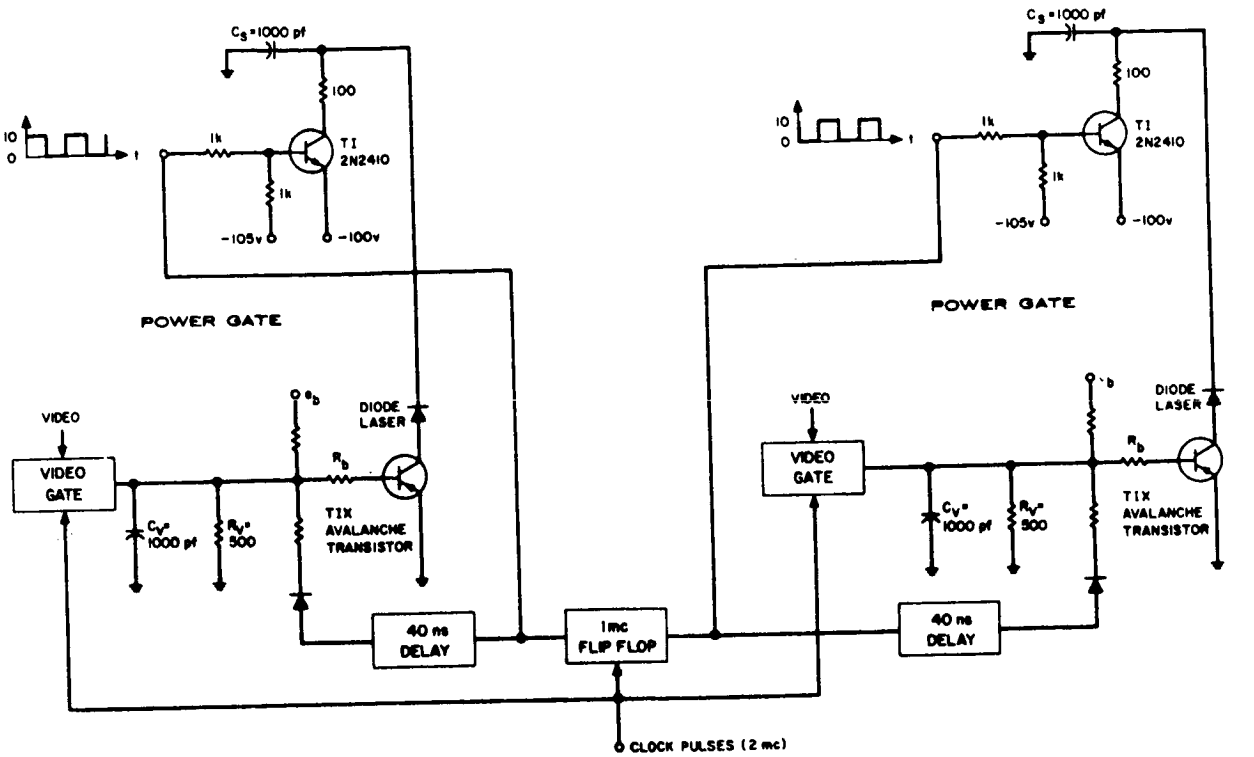
The signal gate shown in Figure A6-5, with a switching level stable to 0.002 volt for the avalanche transistor, will require a maximum video signal of only 0.2 volt, which may be gated into capacitor C_V in a matter of 40 nanoseconds. This gate is the only unavoidable source of deadtime in the system and, therefore, efforts must be made to minimize its switching time.

The inductance in the storage capacitor must be limited to very small levels to achieve proper circuit operation for small storage capacitors of 1000 pf or less. Similarly, the lead inductance in the discharge circuit must be held to extremely low values. Calculations indicate that to allow 800 amperes with a rise time of 1 nanosecond in ordinary wire, only 10 microns of wire per volt will be allowable. Thus, special packaging and the use of coaxial cable will be necessary for the discharge circuit.

SUMMARY

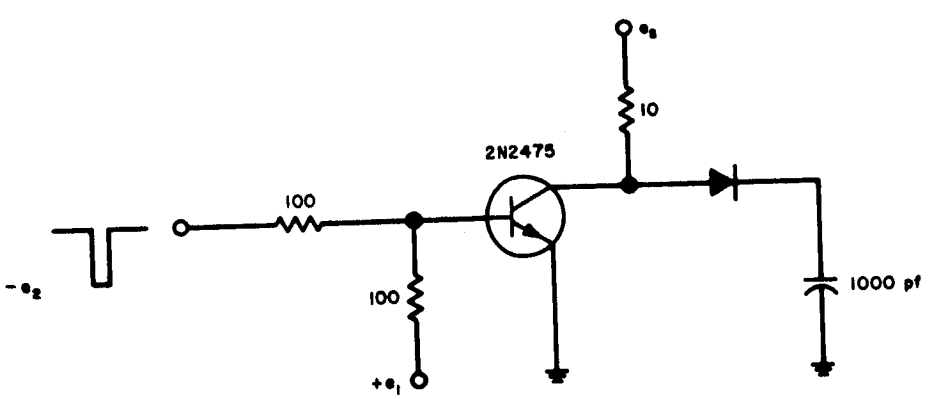
1) Laser diode junction areas should be wider by a factor of 30 to carry 800 amperes in a 1 percent duty cycle for PPM. Net efficiency of the diode laser itself would be 6 percent, assuming 100 percent quantum efficiency. Assuming there is a wider junction, a cooling system is required to conduct 200 watts of power away from the bulk resistance of the diode laser.

2) For linear or logarithmic compression in a PPM system, a threshold switch is required. For the example system, 800 amperes are supplied by this switch with a rise time of 1 nanosecond. At the moment, the only known device to come close to the requirements is an avalanche transistor. However, the best available avalanche transistor for the PPM system must dissipate at least 20 times the power output of the diode laser. It will reduce the efficiency of the system by at least a factor of 20. Transistor technology needs improvement with respect to higher power capabilities and lower avalanche voltages if the avalanche switch is to be used in producing an average laser power output of 12 watts.



POWER GATE DELAY TIME 40 ns
VIDEO GATE CHARGING TIME 40 ns
EACH AVALANCHE TRANSISTOR ALTERNATELY BACK-BIASED BY THE FLIP-FLOP
FOR SYNCHRONIZATION, 40 ns DELAY NEEDED FOR BIAS

Figure A6-4. Laser Injection Diode Modulator Power Gate



MAXIMUM e_s = 0.2v (VIDEO SIGNAL)
MAXIMUM $i_c \simeq$ 20 ma (CHARGING CURRENT)
RC CHARGE TIME $\simeq$ 10 ns
SWITCHING TIME FOR 2N2475 AT 100 ma = 40 ns

Figure A6-5. Laser Injection Diode Modulation Signal Gate

3) High-current, high-speed switching is required to supply the storage capacitor in the system. Transistors most nearly meet these specifications; however, unless two systems are used alternately to allow for the limitations of the transistor, a 50 percent deadtime must be tolerated which is involved in charging the capacitor. For this type of operation additional optical equipment will be needed to position the laser beams.

4) Presently, a deadtime of 40 nanoseconds must be allowed for the switching time of the signal gate transistor; this is assuming a switching stability of 0.002 volt for the avalanche transistor. The result of the deadtime is to further increase the system bandwidth and switching speeds if the system information rate is to be maintained at the same level as for a system without deadtime.

5) Careful consideration must be given to the discharge circuit with respect to packaging since even small values of lead inductance, and inductance in the storage capacitor will degrade the operation of the system.

6) With the alternate system, an avalanche switch current of 15 amperes, a pulse width of 10 nanoseconds (1 percent duty cycle), a deadtime of 40 nanoseconds, a 1/2 watt transmitter power output may be achieved. To increase the system power output some consideration should be given to the possibility of placing several avalanche switches in parallel, with respect to both the input signal and the supply. This may alleviate the current limitation of the switch.

7. DEMODULATION TECHNIQUES

Demodulation, decoding, and detection are used here synonymously to denote the recovering of the information encoded in the electromagnetic field propagating from the transmitting to the receiving apparatus. In this context many devices known as "detectors" are not really detectors but perform some "predetection" function. In a superheterodyne receiver the so-called "first detector" acts as an amplifier and frequency translator while the real detection takes place in the "second detector." On the other hand, in a video receiver, detection takes place immediately in the first stage, while a bandpass optical filter may or may not be employed in front of it.

The optical receiver not only uses components that differ from those in the video receiver and performsfunctions similar to those performed by the corresponding components in a microwave receiver, but its ultimum detection sensitivity is governed by the uncertainty principle.

Ultimate detection sensitivity has been discussed in the literature (References 7-1 through 7-5) but a clarification of the problem is still desirable. The noise is described in Section 8. Here, the various types of detection will be discussed and their principles of operation identified, particularly the power signal-to-noise ratio (S/N) for various conditions of noise. Then the ultimate detection sensitivity which may be obtained from each will be examined and the advantages and disadvantages of each outlined.

VIDEO DETECTION

The operation of a direct detection receiver is illustrated in Figure 7-1. The incoming signal of average power $\bar{P}_s$ of the receiver passes through an optical filter of bandwidth B_i cps and impinges on a photosensitive surface resulting in a current output:

$$i_s = \eta q \bar{n}_s = \eta q \frac{\bar{P}_s}{h\nu} \qquad (7\text{-}1)$$

where

η = quantum efficiency of the photosensitive surface

q = charge of an electron

ν = frequency of impinging photon

$\bar{n}_s = \frac{\bar{P}_s}{h\nu}$ = number of photons of frequency ν per second composing the signal

The signal power in the output, referred to unity resistance for simplicity, is:

$$S = i_s^2 = (\eta q \bar{n}_s)^2 \tag{7-2}$$

The noise is composed of:

1) Background noise caused by unwanted sources radiating at the frequencies of the passband of the input optical filter. These are considered usually as blackbody radiation at a temperature characterizing the background. For the frequencies under consideration, only temperatures of the order $T >> h\nu/k$ make appreciable contributions to the noise. In general, for a background at temperature T_b the background incident power on the photodetector with an input filter of bandwidth B_i is:

$$\bar{P}_b = k T_b B_i \tag{7-3}$$

This is equivalent to an average number of background photons

$$\bar{n}_b = \frac{\bar{P}_b}{h\nu} = \frac{k T_b}{h\nu} B_i \tag{7-4}$$

which give rise to a background current

$$i_b = \eta q \frac{k T_b}{h\nu} B_i \tag{7-5}$$

2) Thermal noise due to the resistive element of the detector

$$N_T = 4k T B_o \tag{7-6}$$

where B_o is the bandwidth of the components after the detector, and R has been taken equal to unity.

3) Shot noise caused by dark current $\bar{i}_d$ in the detector, for frequencies less than the reciprocal of the transient time of an electron from the emitter to the anode

$$N_d = 2q\bar{i}_d B_o \tag{7-7}$$

4) Quantum noise due to signal and background incident power on the photodetector. Because there is no correlation between signal and background photons, the total number of photons incident on the photodetector is $(\bar{n}_s + \bar{n}_b)$ resulting in an average current in the detector's output:

$$\bar{i} = \eta q (\bar{n}_s + \bar{n}_b) \tag{7-8}$$

This average current results in a shot noise, called quantum noise

$$N_q = 2q\bar{i} B_o = 2\eta q^2 (\bar{n}_s + \bar{n}_b) B_o \tag{7-9}$$

The total noise is

$$N = N_T + N_d + N_q \tag{7-10}$$

and the signal-to-noise ratio,

$$\frac{S}{N} = \frac{(\eta q \bar{n}_s)^2}{[4kT + 2q(\bar{i}_d + \eta q(\bar{n}_s + \bar{n}_b)]\ B_o} \tag{7-11}$$

Here the dark current may be referred as equivalent photons in the input of the photodetector,

$$\bar{n}_d = \frac{\bar{i}_d}{\eta q}$$

and obtain

$$\frac{S}{N} = \frac{(\eta q \bar{n}_s)^2}{[4kT + 2\eta q^2(\bar{n}_s + \bar{n}_b + \bar{n}_d)]B_o} \tag{7-12}$$

As already mentioned, the effect of the thermal noise is negligible; therefore, $4kT$ may be neglected. Then

$$\frac{S}{N} = \frac{\eta \bar{n}_s^2}{2(\bar{n}_s + \bar{n}_b + \bar{n}_d) B_o} \tag{7-13}$$

If the number of background and dark current photons is assumed to be very small in comparison to the number of signal photons, Equation 7-13 may be modified to give

$$\frac{S}{N} = \frac{\eta \bar{n}_s}{2B_o} \tag{7-14}$$

HETERODYNE CONVERTER

In the heterodyne receiver (Figure 7-2) the incoming signal is of the same average power $\bar{P}_s$ at the receiver, and it is characterized by an angular frequency

$$\omega_s = 2\pi\nu$$

The instantaneous power at the receiver, provided that the signal is sinusoidal, of the above frequency is

$$P_s = A_s^2 \cos^2 \omega_s t \tag{7-15}$$

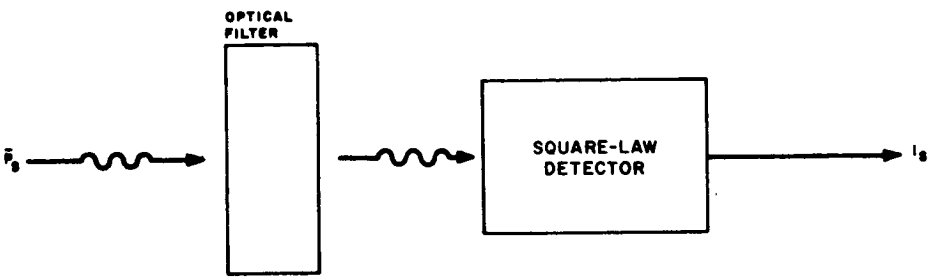


Figure 7-1. Video Detector

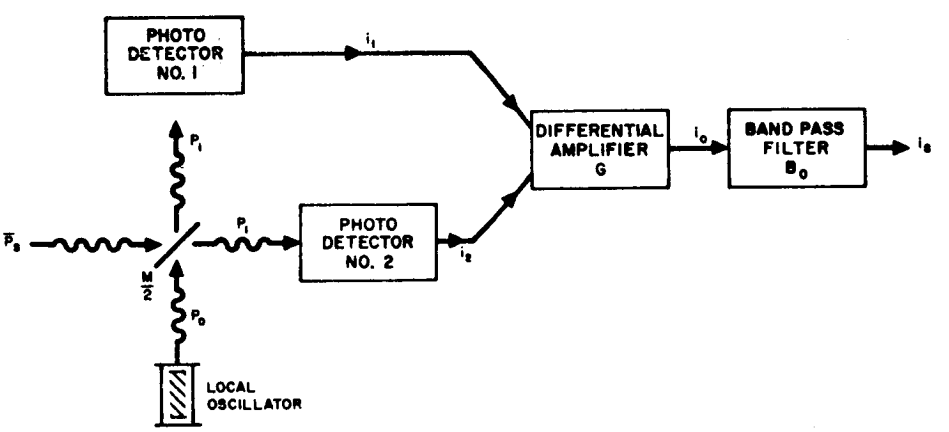


Figure 7-2. Heterodyne Convertor

where A_s is the amplitude of the electromagnetic field. Let the instantaneous power of the local oscillator laser, operating at an angular frequency $\omega_o = 2\pi\nu_o$, be

$$P_o = A_o^2 \sin^2(\omega_o t + \phi) \tag{7-16}$$*

where A_o is the amplitude of the locally generated electromagnetic field.

The half-transmitting, half-reflecting mirror M/2 has a transmission coefficient $\frac{1}{\sqrt{2}}e^{i\pi/4}$ and a reflection coefficient $\frac{1}{\sqrt{2}}e^{-i\pi/4}$. Thus on the photodetectors No. 1 and No. 2, the incident instantaneous power, respectively, is

$$P_1 = \frac{1}{2}\left[A_s \cos(\omega_s t - \tfrac{\pi}{4}) + A_o \sin(\omega_o t + \phi + \tfrac{\pi}{4})\right]^2$$

$$P_2 = \frac{1}{2}\left[A_s \cos(\omega_s t + \tfrac{\pi}{4}) + A_o \sin(\omega_o t + \phi - \tfrac{\pi}{4})\right]^2 \tag{7-17}$$

If both photodetectors have identical quantum efficiencies η, their outputs will be:

$$i_1 = \frac{\eta q P_1}{h\nu} \quad , \quad i_2 = \frac{\eta q P_2}{h\nu} \tag{7-18}$$

Here it is assumed that ω_o is so close to ω_s that the photon energies of the signal and local oscillator are approximately equal. Now in the differential amplifier, of current amplification G, the difference $i_1 - i_2$ is formed, and in its output is realized a current

$$i_o = G(i_1 - i_2) = G\frac{\eta q}{h\nu}(P_1 - P_2) \tag{7-19}$$

*Here $\sin(\omega_o t + \phi)$ is used instead of $\cos(\omega_o t + \phi)$ to facilitate calculations compensating the phase change of the local oscillator wave due to mirror M/2.

but

$$(P_1 - P_2) = \frac{A_s^2}{2} \sin 2\omega_s t + \frac{A_o^2}{2} \sin 2(\omega_o t + \phi) + A_o A_s \cos\left[(\omega_s - \omega_o) - \phi\right]$$

After the filter, which eliminates all frequency components higher than $\delta = \omega_s - \omega_o$, the signal current

$$i_s = G \frac{\eta q}{h\nu} A_o A_s \cos(\delta t - \phi) \tag{7-20}$$

is obtained. The signal power, referred to a unit resistance load, is

$$S = \overline{i_s^2} = \left(\frac{G\eta q}{h\nu}\right)^2 A_o^2 A_s^2 \overline{\cos^2(\delta t - \phi)} = \left(\frac{G\eta q}{h\nu}\right)^2 \frac{A_o^2 A_s^2}{2} \tag{7-21}$$

provided that $\omega_s - \omega_o$ is much larger than the bandwidth $2\pi B_o$ of the circuits following the mixer.

The average power of the signal is

$$\bar{P}_s = \frac{A_s^2}{2} \tag{7-22}$$

and the average power of the local oscillator is

$$\bar{P}_o = \frac{A_o^2}{2} \tag{7-23}$$

Hence the expression for the signal may be written as

$$S = 2\left(\frac{G\eta q}{h\nu}\right)^2 \bar{P}_s \bar{P}_o \tag{7-24}$$

If, instead of the signal in the input, there were background noises of power spectral density $\psi(\nu)$, only components within a bandwidth B_o centered at

the local oscillator frequency ν_o would affect the output (after the bandpass filter). Hence only a background power

$$\bar{P}_b = \psi(\nu)\, B_o \tag{7-25}$$

where

$$\psi(\nu) = \left[\exp\left(\frac{k T_b}{h\nu}\right) - 1\right]^{-1}$$

will be considered as entering the detector and will result. according to the derivation above, to a background noise power

$$N_b = 2\left(\frac{G\eta q}{h\nu}\right)^2 \bar{P}_b\, \bar{P}_o \tag{7-26}$$

The average current out of each photodetector that can be attributed to incident signal and local oscillator intensity is

$$\bar{i}_1 = \frac{\eta q}{h\nu}\,\bar{P}_1 = \frac{\eta q}{4h\nu}\left(A_s^2 + A_o^2\right) = \frac{\eta q}{2h\nu}\left(\bar{P}_s + \bar{P}_o\right) = \bar{i}_2$$

To this may be added the background average incident power $\bar{P}_b$ to obtain the following equation for average current from each photodetector

$$\bar{i} = \frac{\eta q}{2h\nu}\left(\bar{P}_s + \bar{P}_o + \bar{P}_b\right) \tag{7-27}^*$$

This results in shot noise power, referred to a unit resistance load, from each photodetector expressed as

$$N_q = 2q\bar{i}\, B_o = \frac{\eta q^2}{h\nu}\left(\bar{P}_s + \bar{P}_o + \bar{P}_b\right) B_o \tag{7-28}$$

---

*Note that $\bar{P}_b$ here is given by Equation 7-25. It differs from the P_b used in the video detector where the bandwidth used was B_i.

Moreover, if the internal dark current is of an average value $\bar{i}_d$, which may refer to the input at the photodetector as a dark current noise input power $\bar{P}_d = 2h\nu/\eta q \ \bar{i}_d$, it results in dark current shot noise power

$$N_d = 2q\bar{i}_d B_o = \frac{\eta q^2}{h\nu} \bar{P}_d B_o \tag{7-29}$$

The total shot noise for each photodetector is

$$N_s = N_q + N_d = \frac{\eta q^2}{h\nu} (\bar{P}_s + \bar{P}_o + \bar{P}_b + \bar{P}_d) B_o \tag{7-30}$$

Therefore, from Equations 7-26 and 7-30, the total noise in the output of the bandpass filter is

$$N = 2N_s + N_b = \frac{2G^2\eta q^2}{h\nu} (\bar{P}_s + \bar{P}_o + \bar{P}_b + \bar{P}_d) B_o + \frac{2G^2\eta^2 q^2}{h^2\nu^2} \bar{P}_b \bar{P}_o \tag{7-31}$$

From Equations 7-24 and 7-31, the signal-to-noise ratio is

$$\frac{S}{N} = \frac{\eta \bar{P}_s \bar{P}_o}{h\nu B_o (\bar{P}_s + \bar{P}_o + \bar{P}_b + \bar{P}_d) + \eta \bar{P}_b \bar{P}_o} \tag{7-32}$$

When the power of the local oscillator $\bar{P}_o$ is increased the signal-to-noise ratio becomes

$$\frac{S}{N} = \frac{\eta \bar{P}_s}{h\nu B_o + \eta \bar{P}_b} = \frac{\eta \bar{P}_s}{[h\nu + \psi(\nu)] B_o} \tag{7-33}$$

Heterodyne detection eliminates all the noise sources except the quantum noise $h\nu$ and the portion of the background noise which falls within the bandwidth of the output low-pass filter (Equation 7-25). Also, it is easier to construct a bandpass filter in the intermediate frequency $\delta = \omega_s - \omega_o$ than a bandpass filter in the optical frequencies. The bandwidth B_o of this filter

should be sufficient to accommodate all the sidebands created by modulating the transmitting laser, and the doppler shifts and drifts of the transmitting and locally used laser. For low background noise, as it is expected to be, the terms $\eta \bar{P}_b$ may be neglected and

$$\frac{S}{N} = \frac{\eta \bar{P}_s}{h\nu B_o} = \frac{\eta \bar{n}_s}{B_o} \tag{7-34}$$

is obtained. This form of detection yields an improvement of 3 db over video detection (see Equation 7-14).

HOMODYNE DETECTION

The homodyne receiver is diagramed in Figure 7-3. The balanced mixer on the right-hand side is identical to the one described for the heterodyne receiver. Here, however, provisions have been made so that the local oscillator laser operates at the same angular frequency as the incoming signal, i.e., $\omega_o = \omega_s$. Then the signal current (see Equation 7-20) becomes

$$i_s = G \frac{\eta q}{h\nu} A_o A_s \cos \phi \tag{7-35}^*$$

and the associated signal power (to unit resistance) is

$$S = \overline{i_s^2} = G^2 \left(\frac{\eta q}{h\nu}\right)^2 A_o^2 A_s^2 \cos^2 \phi = 4\left(\frac{G\eta q}{h\nu}\right)^2 \bar{P}_s \bar{P}_o \cos^2 \phi \tag{7-36}^*$$

The total noise is still given by Equation 7-31. Here, however, the signal current is at the baseband: the bandwidth B_o is centered at zero frequency. Hence a low-pass filter of bandwidth $B_o/2$ is employed. Thus the signal-to-noise ratio after the assumption of a strong local oscillator signal is

$$\frac{S}{N} = \frac{2\eta \bar{P}_s \cos^2 \phi}{h\nu \frac{B_o}{2} + \eta \bar{P}_b} = \frac{2\eta \bar{P}_s \cos^2 \phi}{\left[\frac{h\nu}{2} + \psi(\nu')\right] B_o} \tag{7-37}$$

*In reality ϕ is not the phase difference between the local oscillator and signal because the local oscillator should be $3\pi/4$ lagging in phase in relation to the signal to compensate for the effect of the mirrors in the phase of the transmitted and reflected waves. This constant phase difference is ignored here to facilitate discussion.

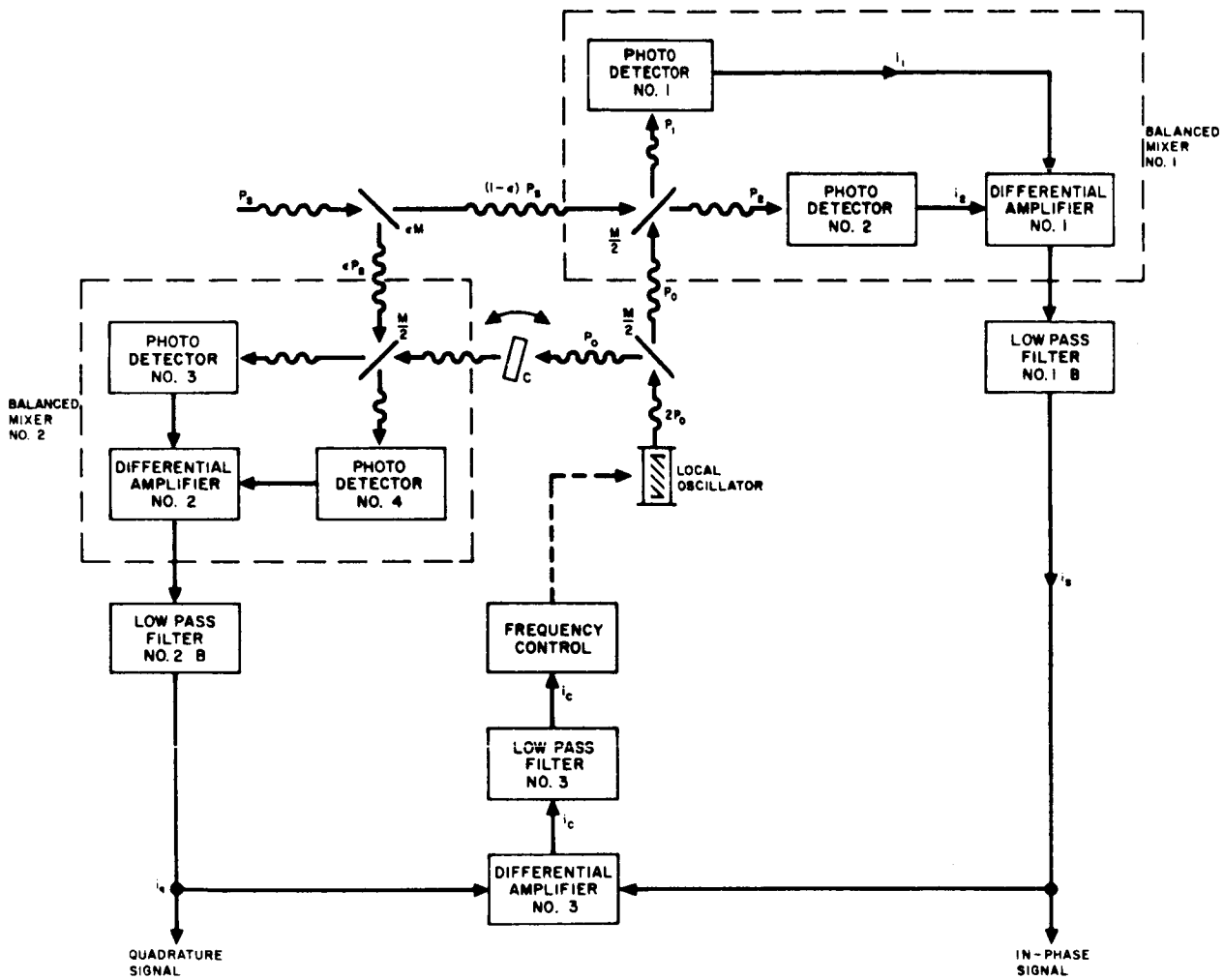


Figure 7-3. Homodyne Detector

Hence, the signal-to-noise ratio is a function of the phase difference ϕ between the incoming signal and the locally generated one. For $\phi = 0$ the maximum signal-to-noise ratio is obtained

$$\frac{S}{N} = \frac{2\eta \bar{P}_s}{\left[\frac{h\nu}{2} + \psi(\nu)\right] B_o} \tag{7-38}$$

If $\psi(\nu) \leq h\nu/2$, as is expected,

$$\frac{S}{N} = \frac{\eta \bar{P}_s}{h\nu \frac{B_o}{2}} = \eta \frac{\bar{n}_s}{\frac{B_o}{2}} \tag{7-39}$$

From the above it follows that it is important to have the local oscillator laser working at the same frequency as the incoming signal, and moreover to synchronize it in phase. To this end a closed-loop phase-lock system is employed, taking advantage of the fact that the information rate necessary to maintain phase lock can be made negligible in comparison to the signaling rate, provided that transmitting and local lasers are of sufficient frequency stability. This may be done, as illustrated in Figure 7-3, by diverting a fraction ϵ of the signal power $\bar{P}_s$ into the auxiliary balanced mixer No. 2. The compensating plate C is rotated, to change the path length from the local oscillator to mixer No. 2, so that when the phase difference ϕ is zero at mixer No. 1 it is $\pm \pi/2$ at mixer No. 2.*

The output of mixer No. 2 after the low-pass filter No. 2, which is identical to the low-pass filter No. 1 with bandwidth B is**

$$i_q = G_2 \frac{\eta q}{h\nu} \epsilon A_o A_s \sin \phi \tag{7-40}$$

This output now is subtracted from i_s (Equation 7-35) in the differential amplifier No. 3. If

$$\frac{G_1}{G_2} = \epsilon \tag{7-41}$$

*See the footnote pertaining to Equations 7-35 and 7-36.

**$B = B_o/2$.

is arranged and unit current amplification factor for amplifier No. 3 is assumed,

$$i_c = \frac{2G_1 \eta q}{h\nu} A_o A_s \sin\left(\frac{\pi}{4} - \phi\right) \qquad (7\text{-}42)$$

is obtained at the output. This current, which depends on ϕ, may be utilized to control the phase of the local oscillator.

The response time of the phase-lock loop is determined by the bandwidth b of the low-pass filter No. 3. If it is assumed that the rate of change of the phase difference ϕ is a fraction ζ of the signaling rate, B, then

$$b = \zeta B$$

may be set. The more stable the lasers employed for transmission and for the local oscillator, the smaller ζ is.

The relation between ζ (measure of relative frequency stability of lasers) and e (fraction of signal power required to extract its phase information) is a matter of further study. That they should be related is evident from the intuitive notion that the larger the instability of the lasers (measured by ζ), the larger the information required to specify their phase relation, and the greater the amount of power (measured by e) that must be utilized.

DETECTION LIMITS

In the detection process a square law video detector may be employed directly, followed by a low-pass filter of bandwidth B_o. At its output a signal-to-noise ratio given by Equation 7-12 is realized. In the presence of a strong interfering background, the number of background photons allowed to reach the detector should be restricted. This can be done by permitting only photons from a very small angle cone containing the signal source to reach the detector (by the entrance optics or more generally by the antenna) and by allowing photons entering from this cone to reach the detector only if they are within a certain energy interval. This energy interval is completely specified by its extreme frequencies as a frequency interval called bandwidth.

The "filter" which is employed in front of the detector may be either completely passive or active. The passive filter allows any photon within its bandwidth to pass unimpeded, while the active filter, sometimes called an amplifier, enhances the photons allowed to pass, giving out more photons than are incident on the filter input. This enhancing can be accomplished in a number of ways:

1) All incident photons within a certain bandwidth may be amplified. In the presence of background photons, this amplifier does not improve the ratio of signal-to-background photons in its passband. As a matter of fact, this ratio is degraded due to the internal noise of the amplifier.

2) All incident photons within a certain bandwidth are not only equally amplified, but also translated in frequency, usually downwards. Also, internal noise problems here degrade the signal-to-noise ratio.

3) The signal photons are amplified more than the background photons by utilizing certain properties of the signal. This kind of amplification, sometimes called "coherent," improves the signal-to-noise ratio.

In principle, cases 1 and 2 do not differ. If it is assumed that the amplifications involved can be done without introduction of internal noise, they should be equivalent to passive filtering.

A closer examination reveals that for the case of linear amplification, in the sense that input quanta are linearly related to output quanta and phase relationship is preserved, the noiseless linear amplifier is physically unrealizable (Reference 7-3) because it violates the uncertainty principle. The result is that even if the thermal noise is ignored, a valid assumption in cooled amplifiers working at very high frequencies, there still remains a source of noise characterized by a spectral density $h\nu$ (Reference 7-4). This is called "quantum noise."

Case 2 does not differ from case 1 in principle. Case 2 clearly refers to heterodyne (frequency translation), and its analysis reveals that it has the same signal-to-noise ratio performance as the ideal linear amplifier (Reference 7-5). On the other hand, it can be seen from Equation 7-34 that for unity quantum efficiency ($\eta = 1$) the noise spectral density is $h\nu$. This is the quantum noise in a perfect heterodyne. The analyses above have been made under the assumption of an ideal detector employed at the output. By ideal detector is meant a conceptual device which performs with the limit of the uncertainty principle, i. e.,

$$\Delta n \, \Delta \phi = \frac{1}{2} \tag{7-43}$$

where Δn is the rms fluctuations in the number of photons and $\Delta \phi$ is the rms phase fluctuation.

Case 3, however, is different in the sense that use is made of something known about, or extracted from, the signal. In the case of the homodyne the discussion between H. A. Haus et al, and B. M. Oliver (Reference 7-6)

reveals that the inherent noise spectrum of the homodyne is $1/2\ h\nu$, provided that detecting either amplitude or phase, but not both, is of interest. Hence in the homodyne the equivalent input quantum noise is only half as large as that occurring in the perfect heterodyne or perfect linear amplifier receiver. Therefore, the detection limit for the homodyne under quantum noise limited operation is half the detection limit of the other two types.

INFORMATION CAPACITY CONSIDERATIONS

The information capacity of an optical channel has never yet been approached. Nevertheless, the capacity indicates the theoretical capability of a system. In addition, capacity may point out the most promising way to arrive at a given information rate level based on the intuitive feeling that it is easier to reach this level in a system with higher capacity than in one with lower capacity.

An optical channel differs from a channel utilizing electromagnetic waves at a lower frequency because quantum effects caused by high frequency are prominent in the former. J. P. Gordon (Reference 7-7) has analyzed such a channel. The main points of his work are summarized in the following paragraphs.

The information capacity, i. e., the maximum rate at which information can be transmitted, of a channel characterized by:

1) Bandwidth B_o, cps

2) Additive white noise of average power $\bar{P}_b$, watts, or spectral density, ψ, i. e., $\bar{P}_b = \psi B_o$

3) The restriction of available average signaling power $\bar{P}_s$, watts is attained when the signal has the characteristics of white noise,

$$C_W = B_o \lg\left(1 + \frac{\bar{P}_s}{\bar{P}_b + h\nu B_o}\right) + \frac{\bar{P}_s + \bar{P}_b}{h\nu} \lg\left(1 + \frac{h\nu B_o}{\bar{P}_s + \bar{P}_b}\right) - \frac{\bar{P}_b}{h\nu} \lg\left(1 + \frac{h\nu B_o}{\bar{P}_b}\right), \frac{\text{bits}}{\text{sec}} \tag{7-44}*$$

This is the capacity of an electromagnetic wave at the point where $\bar{P}_s$ and $\bar{P}_b$ are referred. The additive noise power $\bar{P}_b$ is usually taken as arising from a blackbody at temperature T_n, ^oK, i. e.,

$$\bar{P}_b = h\nu B_o \left[\exp\left(\frac{h\nu}{kTn}\right) - 1\right]^{-1} \tag{7-45}$$

*Here lg denotes logarithm to the base 2.

For the particular case $B_o = 10^9$ cps and $T_n = 290°$ K, the wave capacity is plotted in Figure 7-4 as a function of frequency ν. Equation 7-44 reduces to the classical formula

$$C_w = B_o \lg\left(1 + \frac{\bar{P}_s}{\bar{P}_b}\right) \text{ bits per second} \qquad (7\text{-}46)$$

provided that $\bar{P}_b >> h\nu B$.

Capacity of Channel with Linear Amplifier

A high-gain linear amplifier characterized by the bandwidth B_o is employed at the receiving end, and the effective input noise, where $K \geq 1$, is

$$N_{eff} = K h \nu B_o$$

In the case of a perfect linear amplifier, the term K equals unity. After the amplifier, the channel is characterized by the bandwidth B_o, noise average power $G(\bar{P}_b + N_{eff}) >> h\nu B_o$, and signal power $G\bar{P}_s$, where G is the amplifier power gain.* Hence, the classical formula is applicable and

$$C_{AMPL} = B_o \lg\left(1 + \frac{\bar{P}_s}{\bar{P}_b + K h \nu B_o}\right) \text{bits per second} \qquad (7\text{-}47)$$

Capacity of Channel with Heterodyne Receiver

The signal-to-noise ratio performance of the heterodyne converter is given by Equation 7-33. Here again, for the reasons given above, the classical formula is applicable and

$$C_{HET} = B_o \lg\left(1 + \frac{\bar{P}_s}{\bar{P}_b + \frac{1}{\eta} h \nu B_o}\right) \text{bits per second}$$

is obtained. Comparison of Equations 7-47 and 7-48 reveals that the information capacity of a system using a heterodyne receiver is the same as that of one using a linear amplifier with $K = 1/\eta$.

*The assumption is that after much amplification the information remaining in the wave can be extracted. Therefore, Equation 7-46 characterizes the system using such an amplifier at the front end of its receiver.

Capacity of Channel using Homodyne Receiver

The signal-to-noise ratio performance of the homodyne is given by Equation 7-37. Based on the discussion given above, the classical capacity formula is applicable and hence

$$C_{HOM} = \frac{B_o}{2} \lg \left(1 + \frac{2 \bar{P}_s}{\bar{P}_b + \frac{1}{2\eta} h \nu B_o} \right) \text{ bits per second} \qquad (7\text{-}49)$$

Capacity of Channel with Energy Receiver

This is the system which employs a video detector (quantum counter) directly for detection. It has been analyzed for the case $\bar{P}_s >> h \nu B$ and under this condition its capacity found to be:

$$C_v = \frac{1}{2} C_w - B_o \left[\lg \sqrt{2\pi} + 0.289 \lg e \right] \text{ bits per second} \qquad (7\text{-}50)$$

Information Extraction Ability of Various Systems

To compare the capacities of the various systems, information extraction ability from an electromagnetic wave is defined as

$$\gamma = \frac{C_x}{C_w}$$

where C_x is the capacity for the system under examination. In Figure 7-5 the ability γ is plotted versus frequency in the systems examined above for the particular case of signal power $\bar{P}_s = 10^{-11}$ watt and bandwidth and incident additive noise $\bar{P}_b$ corresponding to that used in Figure 7-4.

In the same figure for purposes of comparison the information extraction ability of the energy-sensitive receiver is plotted. It is employed here particularly as a binary quantum counter. This system is described and analyzed in Reference 7-7. From Figure 7-4 it is evident that for a given bandwidth B_o and signaling power $\bar{P}_s$, the information capacity of an electromagnetic wave diminishes rather rapidly as its frequency ν increases beyond $\bar{P}_s / h B_o$.

From Figure 7-5 it follows that the efficiency with which a receiver can extract the information incorporated in an electromagnetic wave depends on the ratio $\bar{P}_s / h \nu B$, and various types of receivers are more efficient for various regions of this ratio. Thus, the receiver with a perfect heterodyne converter or a perfect linear amplifier is very efficient for $\bar{P}_s >> h \nu B$,

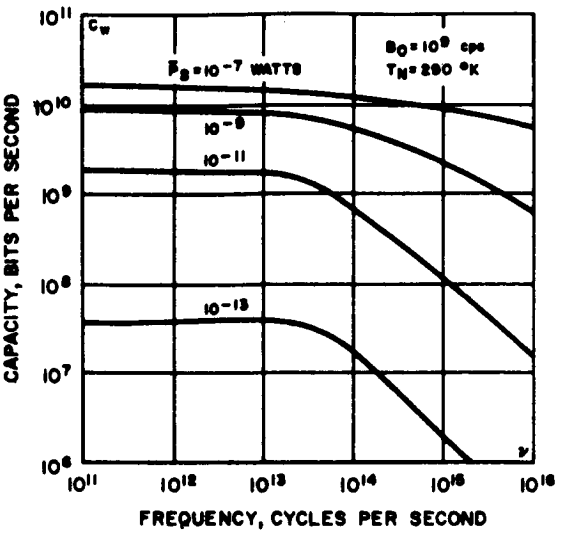


Figure 7-4. Information Capacity of Electromagnetic Wave (Single Mode)

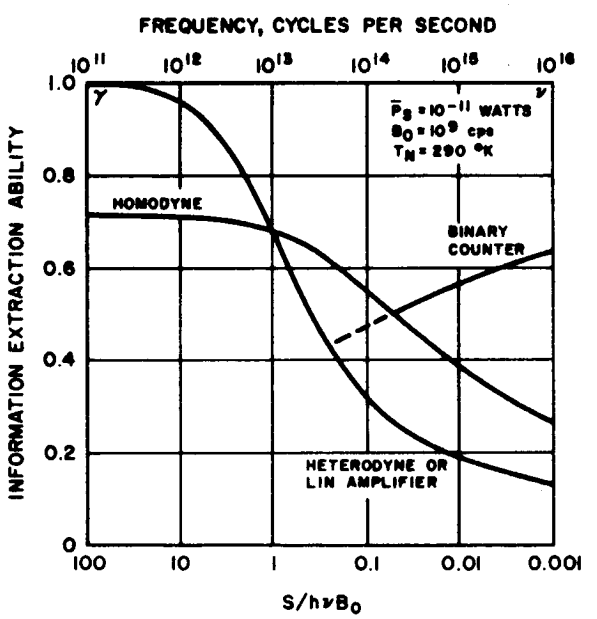


Figure 7-5. Information Extraction Ability for Various Receivers

while the energy-sensitive receiver is half as efficient for the same situation. The homodyne is somewhat between, but assumes predominance in lower values of $\overline{P}_s/h\nu B$ down to the value 0.1 where the binary quantum counter is better, arriving at unity efficiency for a very low signal level.

DISCUSSION

Previously, three detection schemes have been described and analyzed on purely theoretical bases. Now their theoretical capabilities will be summarized and the physical limitations in their implementation discussed.

The video detector can achieve a signal-to-noise ratio at its output equal to

$$\left(\frac{S}{N}\right)_v = \frac{\eta \overline{n}_s^2}{2(\overline{n}_s + \overline{n}_b + \overline{n}_d) B_o} = \frac{\eta \overline{n}_s}{2\left(1 + \frac{\overline{n}_b}{\overline{n}_s} + \frac{\overline{n}_d}{\overline{n}_s}\right) B_o} \tag{7-51}$$

The heterodyne converter exhibits a signal-to-noise ratio performance similar to that of a linear amplifier and equal to

$$\left(\frac{S}{N}\right)_{HET} = \frac{\eta \overline{P}_s}{h\nu B_o + \eta \overline{P}_b} \tag{7-52}$$

The homodyne, on the other hand, has a signal-to-noise ratio of

$$\left(\frac{S}{N}\right)_{HOM} = \frac{2\eta \overline{P}_s}{h\nu \frac{B_o}{2} + \eta \overline{P}_b} \tag{7-53}$$

From the signal-to-noise ratio point of view, the rating is homodyne, heterodyne, video. It is easy to see that the dark current has been eliminated as a source of noise in the homodyne and heterodyne, being wiped out by the local oscillator. Moreover, the background additive noise manifested by $\overline{n}_b$ in the video and $\overline{P}_b$ in the other two systems is not the same. While $\overline{n}_b$ refers to the number of photons entering through the passband B_i of the input optical filter, $\overline{P}_b$ is referred to the bandpass filter in the intermediate frequency with a bandwidth B_o. The construction of narrow-band optical filters is very difficult, and at present it is not possible to obtain an optical filter with a bandpass of less than 1 Å. This is equivalent to a bandwidth of 6.12×10^{10} cps

at a ruby laser midfrequency (0.6943 micron) and 5.4×10^9 cps in the case of a disprosium midfrequency (2.36 microns), while the modulated signal may occupy a band of less than 10^7 cps. Therefore a reduction of background noise of the order of 10^3 is obtained by the employment of a more easily constructed filter of a bandwidth 10^7 cps centered at the intermediate frequency.

That the signal-to-noise ratio performance of the homodyne is at least 3 db better than that of the heterodyne is evident from Equations 7-52 and 7-53. On the other hand, even with the absence of dark current and background noise, the heterodyne has a signal-to-noise ratio performance 3 db above the video. This last figure, however, assumes that an ideal detector is employed in the output of the heterodyne. If, however, this ideal detector is substituted by a square law device with an efficiency η_2 which performs in a way similar to the video detector, the situation is identical to having in the input of this second detector a signal $\eta \bar{P}_s = \eta \bar{n}_s h\nu$ and a noise $h\nu B_o + \eta \bar{P}_b = h\nu(B_o + \eta \bar{n}_b)$. Therefore, the signal-to-noise ratio realized at its output will be, by virtue of Equation 7-51,

$$\left(\frac{S}{N}\right)_{v_2} = \frac{\eta_2 \eta \bar{n}_s}{2\left(1 + \frac{B_o}{\eta \bar{n}_s} + \frac{\bar{n}_b}{\bar{n}_s} + \frac{\bar{n}_{d_2}}{\eta \bar{n}_s}\right)B_o} \tag{7-54}$$

where $\bar{n}_{d_2}$ are the dark current noise photons in the second detector. Now assuming the original background to be negligible ($\bar{n}_b = 0$), the ratio of the signal-to-noise ratio at the output of a video detector (Equation 7-51) to the signal-to-noise ratio of a video detector employed after the heterodyne converter, is

$$\frac{\left(\frac{S}{N}\right)_v}{\left(\frac{S}{N}\right)_{v_2}} = \frac{1}{\eta_2} \frac{\left(1 + \frac{B_o}{\eta \bar{n}_s} + \frac{\bar{n}_{d_2}}{\eta \bar{n}_s}\right)}{\left(1 + \frac{\bar{n}_d}{\bar{n}_s}\right)} \tag{7-55}$$

It follows that heterodyne is profitably employed only if

$$\eta_2\left(1 + \frac{\bar{n}_d}{\bar{n}_s}\right) > 1 + \frac{B_o}{\eta \bar{n}_s} + \frac{\bar{n}_{d_2}}{\eta \bar{n}_s} \tag{7-56}$$

The efficiency η_2 of a square law detector of the intermediate frequency may approach unity while its dark current noise may be considered negligible (by cooling) when Equation 7-56 reduces to

$$\bar{n}_d > \frac{B_o}{\eta}$$

or

$$\bar{n}_d \, h\nu = N_d > \frac{h\nu}{\eta} B_o \tag{7-57}$$

Hence, the employment of the heterodyne is profitable only if the total dark current noise power is larger than the total quantum noise $h\nu/\eta \; B_o$.

The optical heterodyne exhibits the attribute of angular selectivity. If heterodyne action is to be obtained, the signal and local oscillator light beams must be parallel, coherent, and of the same polarization at the photosurface. Otherwise, the resulting intermediate frequency signals will have different relative phases and will tend to cancel out. The beams must be parallel to within

$$\Delta\theta = \frac{\lambda}{d} \tag{7-58}$$

where λ is the optical wavelength and d is the diameter of the photodetector surface (References 7 and 8). This results in a very small angle (for λ = 0.5 micron and the photocathode diameter equal to 1 cm, $\Delta\theta$ is 10 seconds of arc). Precise optical alignment is then required. This increases the complexity of the system and presents serious aiming problems. The effect of such directivity on the receiver signal-to-noise ratio is not yet clear.

The coherency of the light beams to be mixed was postulated in the analysis. Lasers are capable of producing coherent light, but the effect of a turbulent atmosphere on the coherent radiation coming from the deep space vehicle is a problem that has not been resolved. The limits of coherency degradation for profitable employment of mixing is another area for theoretical and experimental study (Reference 7-9).

The local oscillator laser provides the frequency reference in the heterodyne with which to compare the modulated incoming wave. Its stability, therefore, is most important; however, the tolerable limits of this stability remain as a subject for further study. Problems associated with the phase-lock loop of the homodyne cannot be anticipated until a thorough analytical design is completed. The important points of the present investigation have been summarized in Table 7-1. Other demodulation methods such as electro-optical heterodyne and demodulation of polarization modulation were reported previously on pages 63 and 68, respectively, of the Third Interim Report.

TABLE 7-1. TYPES OF DEMODULATION SYSTEMS

	Advantages	Disadvantages	Areas for Further Study
Video detection	1) Simple straightforward system. 2) Subcarrier demodulation is possible. 3) For low $\bar{P}_s/h\nu B_o$ ratio exhibits the best ability to extract information from an electromagnetic wave, used as binary counter.	1) Filtering of background noise takes place at optical frequencies. 2) Has lowest signal-to-noise ratio as compared to other systems. 3) Cannot be used to detect frequency modulation.	1) Employment of a video detector as a binary quantum counter in the case of $\bar{P}_s/h\nu B_o$.
Heterodyne	1) Improved signal-to-noise ratio performance. a) Dark current noise is suppressed. b) Permits filtering of background noise at intermediate frequency. 2) Angular discrimination. Very narrow beam receiver. 3) Heterodyne amplification takes place since signal power is proportional to local oscillator intensity.	1) Increased system complexity. 2) Requires stable local oscillator and alignment between local oscillator and signal.	1) Stability of local oscillator and its control. 2) Optical alignment and its effects on the receiver S/N. 3) Effects of a turbulent medium (atmosphere) on the coherency of a light beam; degree of coherence required for mixing. 4) Effects of statistical fluctuation of quantum efficiencies of the mixing photosurface on the performance of the heterodyne.

TABLE 7-1 (continued)

	Advantages	Disadvantages	Areas for Further Study
Heterodyne (con't.)	4) Best ability to extract information from electromagnetic wave, when $\bar{P}_s/h\nu B$ is larger than 10. 5) Suitable for frequency modulation detection.		
Homodyne	1) Best signal-to-noise ratio performance. 2) Homodyne conversion gain. 3) The best system to extract information from an electromagnetic wave for $0.1 < \bar{P}_s/h \quad B <$ 10 is necessary. 4) Suitable for frequency modulation detection.	1) Same as in Heterdyne. 2) Very stable local oscillator is required. 3) Phase synchronization.	1) Same as in heterodyne, plus the problem associated with the phase-lock loop. 2) What is the minimum number of photon to be received so phase can be defined.

REFERENCES

7-1. H. Friedburg, "General Amplifier Noise Limits," in Quantum Electronics, C.H. Townes, Ed., Columbia University Press, N.Y. (1960), pp 228.

7-2. R. Serber and C.H. Townes, "Amplification and Complementarity," Quantum Electronics (1960), pp 233.

7-3. H. Heffner, "Fundamental Noise Limit of Linear Amplifiers," Proc. IRE 50: 1604 (1962).

7-4. M.W.P. Strandberg, "Inherent Noise of Quantum-mechanical Amplifiers," Phys. Rev. 106: 617 (1957).

7-5. B.M. Oliver, "Signal-to-Noise Ratios in Photoelectric Mixing," Proc. IRE, 49: 1960 (1961).

7-6. H.A. Haus and C.H. Townes, "Comments on Noise in Photoelectric Mixing," Proc. IRE 50: 1544 (1962).

7-7. J.P. Gordon, "Quantum Effects in Communications Systems," Proc. IRE 50: 1898 (1962).

7-8. A.T. Forester, "Photoelectric Mixing as Spectroscopic Tool," J. Opt. Soc. Am. 51: 253 (1961).

7-9. A.T. Forester, R.A. Gudmunsen, and P.O. Johnson, "Photoelectric Mixing of Incoherent Light," Phys.Rev. 99: 1961 (1955).

8. CHANNEL CHARACTERIZATION

An information channel is characterized by a bandwidth occupancy and the unwanted disturbances interferring with the information content of the signal transmitted from the source to destination. For the unwanted interference the general term "noise" is used. The intervening noise sources which may be encountered in a Deep Space Communication system may be classified in the following way:

1) Thermal background, due to stellar and planetary sources.

2) Attenuation, absorption, and scattering noise which results from propagation through the atmosphere.

3) Scattering and absorption noise which results in the light-gathering process.

4) Various inherent and internal noise sources which emerge during the conversion of the optical signal into an electrical waveform.

5) Electrical noise which occurs during the amplification and electrical signal processing, etc.

If all the noise sources could be statistically described, the channel would be completely characterized and, in principle, the method of optimal utilization of the channel could be specified. Such a task is beyond the scope of this study. An attempt is made here to give as complete a characterization of the noise sources as is reasonably commensurate with the goals of the overall analysis. First, however, the attenuation suffered by the signal will be studied. This is because the final utilization of the transmitted signal is made at the detector's output, where the attenuated and transduced version of the original signal is extracted from the noise.

ATTENUATION EFFECTS

Consider a diffraction-limited antenna transmitting an average power $\bar{P}_t$ into a conical beam of angle θ_t, called "beamwidth" at a radiating

wavelength λ (or frequency $\nu = c/\lambda$). At a distance R the power $\bar{P}_t$ has been spread out over an area $(\pi/4)(\theta_t R)^2$, provided that θ_t is very small (so that the approximation $\sin(\theta_t/2) = \theta_t/2$ is valid). Now under the assumption* that $\bar{P}_t$ is uniformly distributed over the area $(\pi/4)(\theta_t R)^2$, a receiving antenna of aperture area A_r placed perpendicular to the radiation at the distance R intercepts power $\bar{P}_r$, such that

$$\bar{P}_r = \bar{P}_t \frac{4 A_r}{(\theta_t R)^2} \tag{8-1}$$

If there are path components of transmittances $\tau_{p_1}, \tau_{p_2} \ldots \tau_{p_1}$ between the two antennas, Equation 8-1 should be modified to:

$$\bar{P}_r = \tau_p \bar{P}_t \frac{4 A_r}{(\theta_t R)^2}$$

where $\tau_p = \tau_{p_1} \tau_{p_2} \ldots \tau_{p_n}$ is the total transmittance of the path between the antennas. For a circular aperture of diameter d_r, the last equation is further reduced to:

$$\bar{P}_r = \tau_p \bar{P}_t \left(\frac{d_r}{\theta_t R}\right)^2$$

The transmitting power $\bar{P}_t$ comes from a source of electromagnetic radiation (in this case from a laser). If the transmittance of the path from the output of the laser to the output of the antenna is τ_t, termed "transmittance of the transmitter," the power generated by the laser $\bar{P}_L$ should be

$$\bar{P}_L = \frac{\bar{P}_t}{\tau_t}$$

---

*This assumption suggests that the beamwidth θ_t should be taken as the angle between the half-power points in the antenna transmission pattern, i.e., $\theta_t = 1.02 \lambda/d_t$ (Reference 8-1) instead of the more conventional diffraction limit value $\theta_t = 1.22 \lambda/d_t$ (Reference 8-2).

The power received $\bar{P}_r$ is utilized at the detector. If the optical path between input to the receiving antenna and input to the detector has a transmittance τ_r, called "transmittance of the receiver," the power $\bar{P}_s$ incident on the detector is

$$\bar{P}_s = \tau_r \bar{P}_r$$

Finally, the relation between laser output power and power input to the detector is:

$$\bar{P}_s = \bar{P}_L \frac{d_r^2}{(\theta_t R)^2} \tau_r \tau_p \tau_t \qquad (8\text{-}2)$$

For the communication links under consideration here, the only known important component which may interfere in the propagation of the electromagnetic field between the two antennas is the atmosphere (cases of earth-to-space links, as well as ground links). The total path transmittance, therefore, is the atmospheric transmittance, i.e., $\tau_p = \tau_a$, which is given as a function of wavelength in Figure 8-1.

Other possible path components may be clouds in the atmosphere, dust, etc., and exoterrestial path components, i.e., the Van Allen belts, meteorites, cosmic dust, etc. The transmittance of the exoterrestial path component is taken to be unity in the absence of better information. The path component in the atmosphere (clouds, dust, etc.) has been discussed in previous reports, and the difficulties of establishing an absolute, simple characterization is indicated. For present purposes, however, it is assumed that these effects are characterized by the data in Figure 8-1.

Moreover, the transmittances τ_t and τ_r of the optical paths in the transmitter and in the receiver are assumed to be given by the constant values

$$\tau_t = 0.85$$

$$\tau_r = 0.65$$

To a certain extent, control may be exercised in the construction of the path components by choosing the best materials available at the operational wavelength (see Figure 8-2). Typical path components may include the lenses, optical filters in the receiver, modulation cells in the transmitter, etc. The values chosen above represent reasonable values for the optical frequencies considered. From Equation 8-2 it is easy, now, to find the signal current output of the detector. Clearly $\bar{P}_s/h\nu$ is the number of photons per second

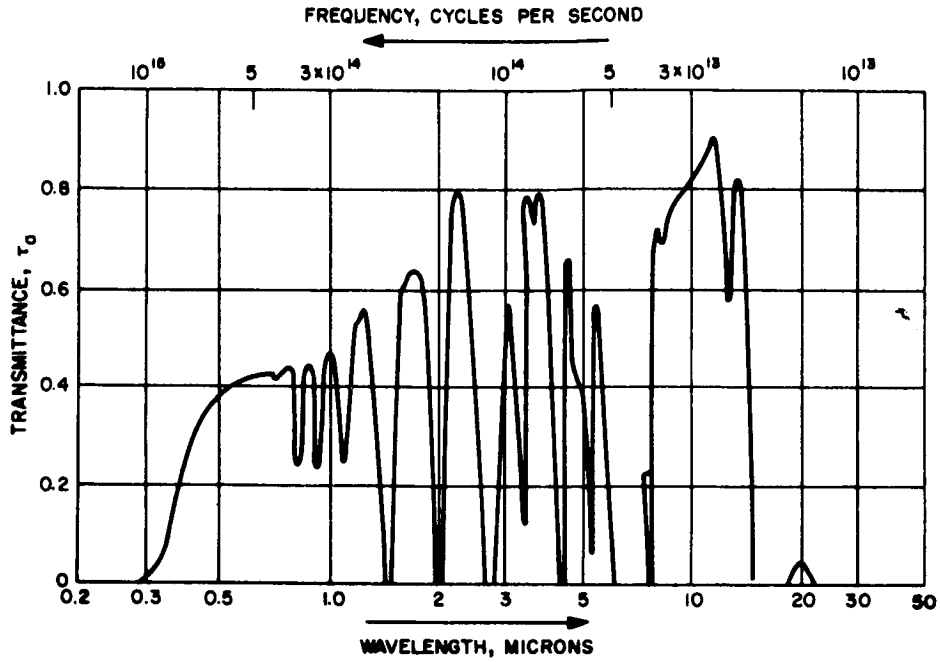
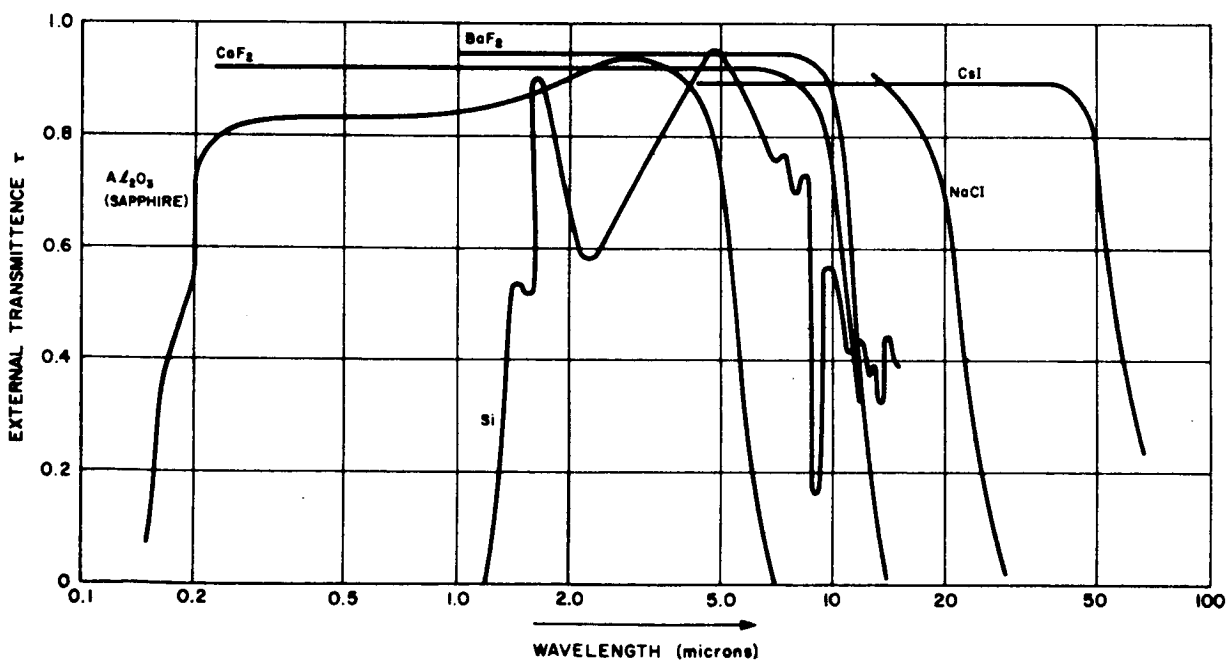


Figure 8-1. Atmospheric Transmittance, τ_a



SOURCE: UNIVERSITY OF MICHIGAN, INSTITUTE OF SCIENCE AND TECHNOLOGY. INVESTIGATION OF OPTICAL SPECTRAL REGIONS FOR SPACE COMMUNICATIONS, 2ND INTERIM REPORT, NOV 1962

Figure 8-2. External Transmittance of Selected Material in the 0.1- to 100-Micron Region

incident on the detector. If the quantum efficiency of the detector is η, at the transmitting frequency considered, in its output are realized $\eta \bar{P}_s/h\nu$ electrons per second (called photoelectrons); hence the current due to the signal photons is:

$$\bar{i}_s = \eta \frac{\bar{P}_s}{h\nu} = \frac{d_r^2 \tau_r \tau_p \tau_t \eta}{(\theta_t R)^2 h\nu} \bar{P}_L \qquad (8\text{-}3)$$

DESCRIPTION OF NOISE SOURCES

The noise entering the communication system under study will be examined at the output of the photodetective apparatus, where the signal has been referred to in the previous paragraph.

It is thus assumed that the noise of the following electronic circuits is either very small or, at any rate, common for all optical communications systems and will not be examined here.

In general, the noise of an optical communication system may be considered to be composed of two main parts:

1) Noise which enters the detector with the signal, described under the general term "background noise N_b."

2) Noise generated in the detector called "detector noise N_d."

There are many ways of describing the noise; however, all of them are underlined, from the statistical communication point of view, by the same principle; namely, that the statistics of the noise are normal (Gaussian) and therefore the power spectral density (i.e., power distribution versus frequency) completely determines the fluctuations due to the noise. Moreover, the physical fact that the blackbody radiation presents a more or less constant spectral density within small frequency intervals (bandwidths) results in the noise examined in this small bandwidth to be uncorrelated (white); i.e., the instantaneous noise amplitude is independent of its value at any previous instant. On the other hand, if in a given bandwidth uncorrelated noise of a certain spectral density level is observed, this noise may be considered as due to a blackbody at such a temperature T_E that the same spectral density level results in the bandwidth under consideration. This temperature T_E is known as the "equivalent noise temperature" of the noise observed in the given bandwidth.

Background Noise

There are many ways of describing the background noise spectrum. Here, however, the notion of the spectral irradiance $H\lambda$ is used, i.e., the

incident power (watts) per unit area (square centimeters) perpendicular to the incoming radiation per unit wavelength (microns).

An antenna with an aperture area A_r will receive noise from a background with irradiance $H_b(\lambda)$

$$P_{b_1} = A_r H_b(\lambda) \Delta\lambda \tag{8-4}$$

where $\Delta\lambda$ is the bandwidth over which the detector is restricted to operate. It is worth noting that the bandwidth $\Delta\lambda$ is expressed here in a wavelength interval and therefore is frequency dependent. Its equivalent bandwidth $\Delta\nu$ in cycles per second is given by:

$$\Delta\nu = \frac{c}{\lambda^2} \Delta\lambda, \qquad \text{cps} \tag{8-5}$$

($c\lambda$ and $\Delta\lambda$ are in the same length units.) For a circular aperture of diameter d_r and a receiver transmittance τ_r, the incident noise power on the detector is

$$P_b = \frac{\pi d_r^2}{4} H_b(\lambda) \Delta\lambda \ \tau_r$$

The resultant noise current output from the detector due to the background noise, for a detector of quantum efficiency η, is

$$\bar{i}_b = \frac{\eta}{h\nu} \frac{\pi d_r^2}{4} H_b(\lambda) \Delta\lambda \ \tau_r \tag{8-6}$$

where $\nu = c/\lambda$ is the midband frequency.

Observe that Equation 8-6 may have different forms for various expressions of the irradiance. Thus, sometimes the irradiance is given at the edge of the earth's atmosphere. Then for an antenna on the earth's surface the irradiance is

$$H_b(\lambda) \ \tau_a(\lambda)$$

where $\tau_a(\lambda)$ is the total transmittance of the atmosphere at the wavelength indicated (see Figure 8-1). In some applications the irradiance is given as

$Q_b(\lambda)$ in terms of photons-cm^{-2}-μ^{-1} instead of watts-cm^{-2}-μ^{-1}. It is useful to know then that:

$$Q_b(\lambda) = \frac{H_b(\lambda)}{h\frac{c}{\lambda}} \tag{8-7}$$

The sources of background noise have been described in previous reports.

The results are summarized and presented in terms of irradiance in Figures 8-3 through 8-7.

Note that the background noise due to stars is neglected because its contribution is negligible.

Also observe that for a ground antenna the daytime solar contribution disappears at night leaving only the earth's contribution (earth at 250 °K).

Detector Noise

The causes of internal noise are many and their mechanism rather involved, and they have been treated in the literature.

Mainly there are:

1) Johnson Noise: Due to the resistive element r of the detector

$$N_T = 4\,k\,t\,B_o\,r \tag{8-8}$$

where B_o is the bandwidth of the circuits measuring the (signal) noise, after the detector, T is the absolute temperature of the detector, and k is the Boltzmann's constant. Usually this noise is eliminated by cooling the detector.

2) Current Noise: The mechanism of this noise remains uncertain. Its accepted value though is given by:

$$N_i = K\frac{\bar{I}_t^{\,\alpha}}{f\,A_d W}\,B_o \tag{8-9}$$

where K is a constant

A_d is the sensitive area of the detector
W is detector's thickness
$\bar{I}_t$ is the total average current through the detector
f is the frequency at which the measurement takes place (modulation frequency)
α is an exponent in the range 1.25 to 4, but usually taken as 2.

Clearly this noise is insignificant for high coding (modulation) rates.

3) Generation, Recombination Noise: This is the characteristic noise in semiconductors. It is caused by the raise of valance-band electrons in the conduction band (this is called generation because it creates mobile carriers for current flow), and also by the recombination of electrons and holes. Its value is:

$$N_g(f) = \frac{4\,\bar{I}_t^{\,2}\,\sigma}{\bar{n}\,(1 + 4\pi^2 f^2 \sigma^2)} \tag{8-10}$$

where $\bar{n}$ is the average number of free electrons
σ is the carrier lifetime

Its value is insignificant as far as $f\sigma >> 1$

4) Shot Noise: Due to the discrete nature of the electric charge (i.e., the electron) and its value is

$$N_s = 2\,q\,\bar{I}_t\,B_o \tag{8-11}$$

where q is the electron charge.

It has already been mentioned that the noise due to sources 1, 2, and 3 can be neglected easily, in principle. This is not so, however, for the shot noise. This noise depends on the total average current $\bar{I}_t$ through the detector, which consists of the average signal current $\bar{i}_s$, the average background noise current $\bar{i}_b$, and the average current which flows through the detector in the absence of any input. This latter current is called average "dark current $\bar{i}_d$" and gives rise to the dark current noise

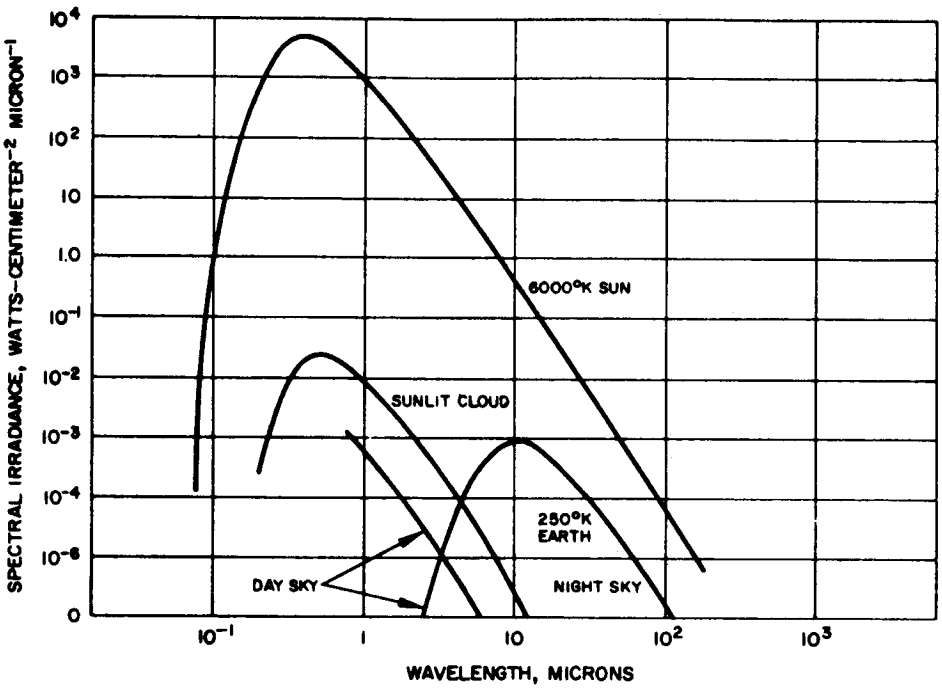
$$N_d = 2\,q\,\bar{i}_d\,B_o \tag{8-12}$$

More often, however, the noise performance of a detector is described in noise equivalent power (NEP) described in Section 2.

REFERENCES

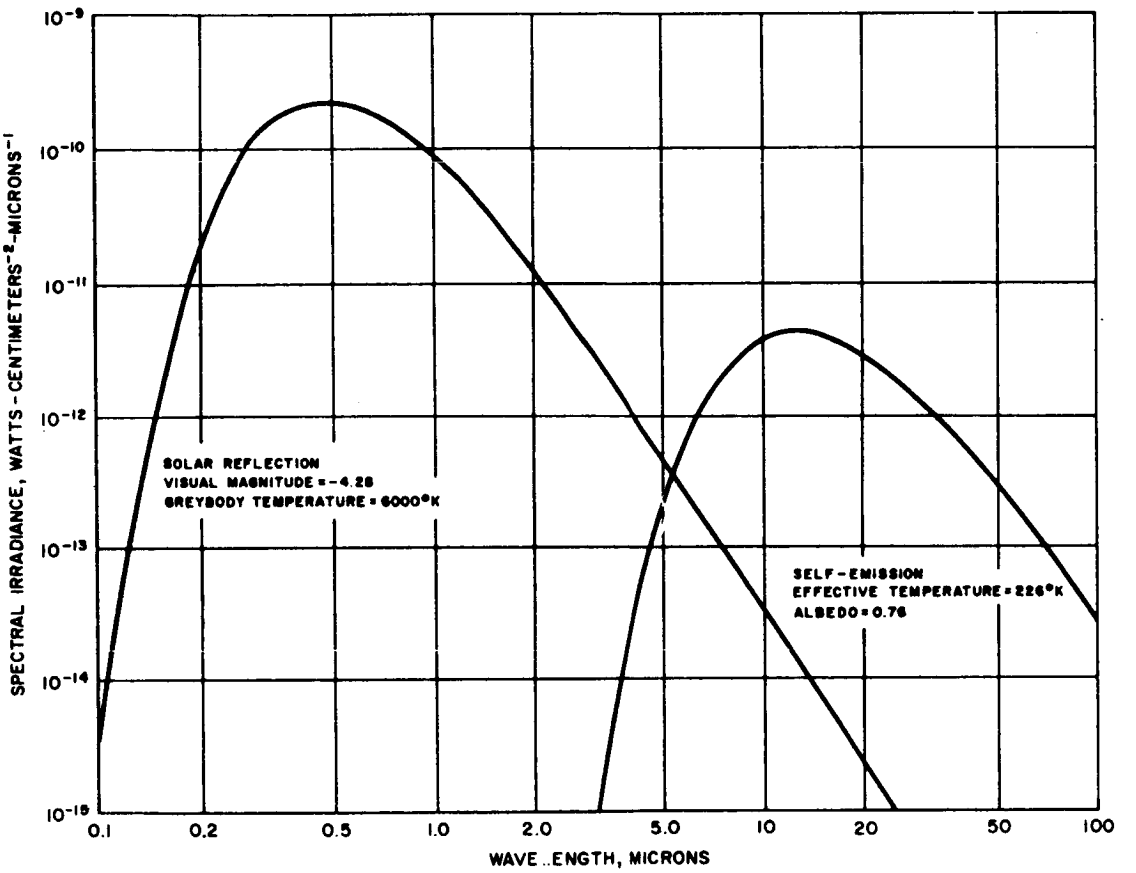
8-1. M. I. Skolnik, Introduction to Radar Systems, McGraw-Hill, N. Y., 1962.

8-2. M. Born and Wolf, Principles of Optics, Pergamon Press, N. Y., 1959.



SOURCE: W E HORN, OPTICS OF THE ATMOSPHERE AND SPACE DELIVERED AT UCLA 1962

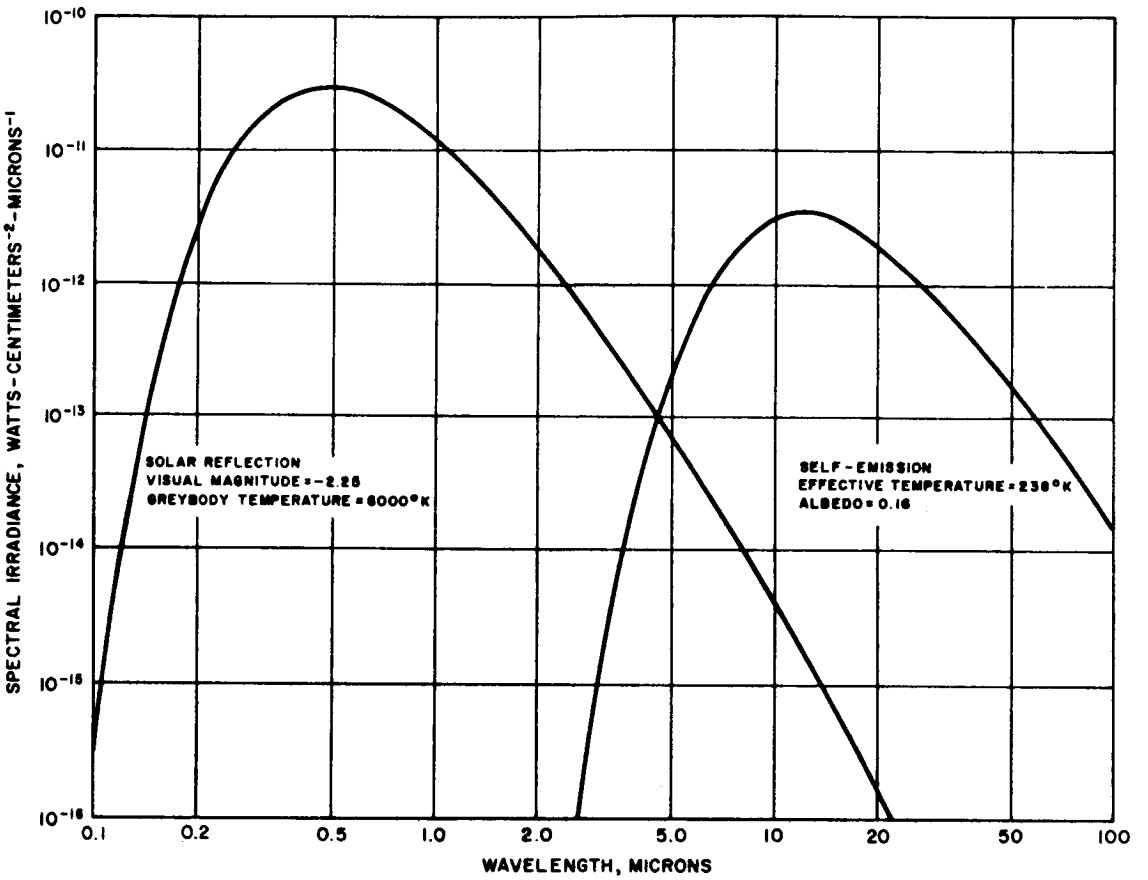
Figure 8-3. Spectral Irradiance of Sun, Earth, Bright Cloud, and Sky on Earth's Surface



SOURCE: UNIVERSITY OF MICHIGAN, INSTITUTE OF SCIENCE AND TECHNOLOGY. INVESTIGATION OF OPTICAL SPECTRAL REGIONS FOR SPACE COMMUNICATION, 2ND INTERIM REPORT, NOV 1962

Figure 8-4. Calculated Maximum Irradiance from Venus Outside Earth's Atmosphere

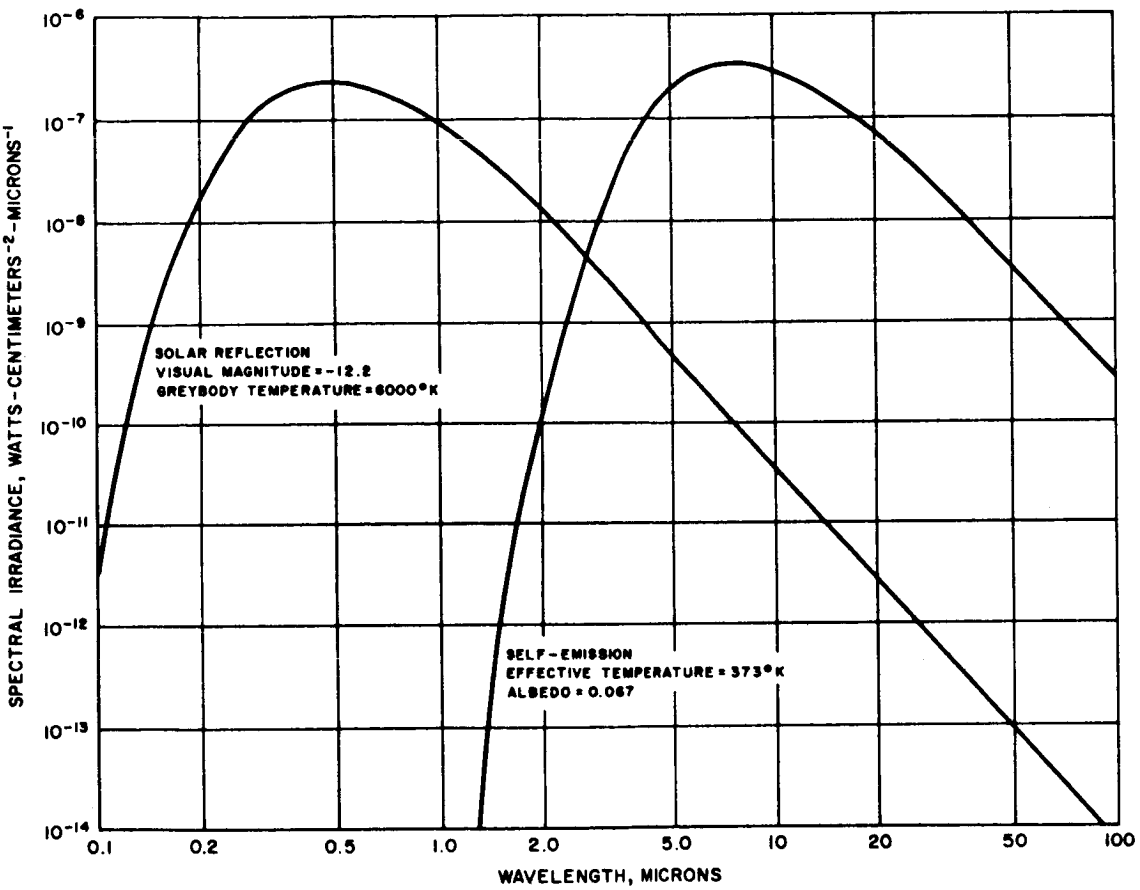
Venus to earth at minimum distance of 10.3(10^3) km



SOURCE: UNIVERSITY OF MICHIGAN, INSTITUTE OF SCIENCE AND TECHNOLOGY. INVESTIGATION OF OPTICAL SPECTRAL REGIONS FOR SPACE COMMUNICATIONS, 2ND INTERIM REPORT, NOV 1962

Figure 8-5. Calculated Maximum Irradiance from Mars Outside Earth's Atmosphere

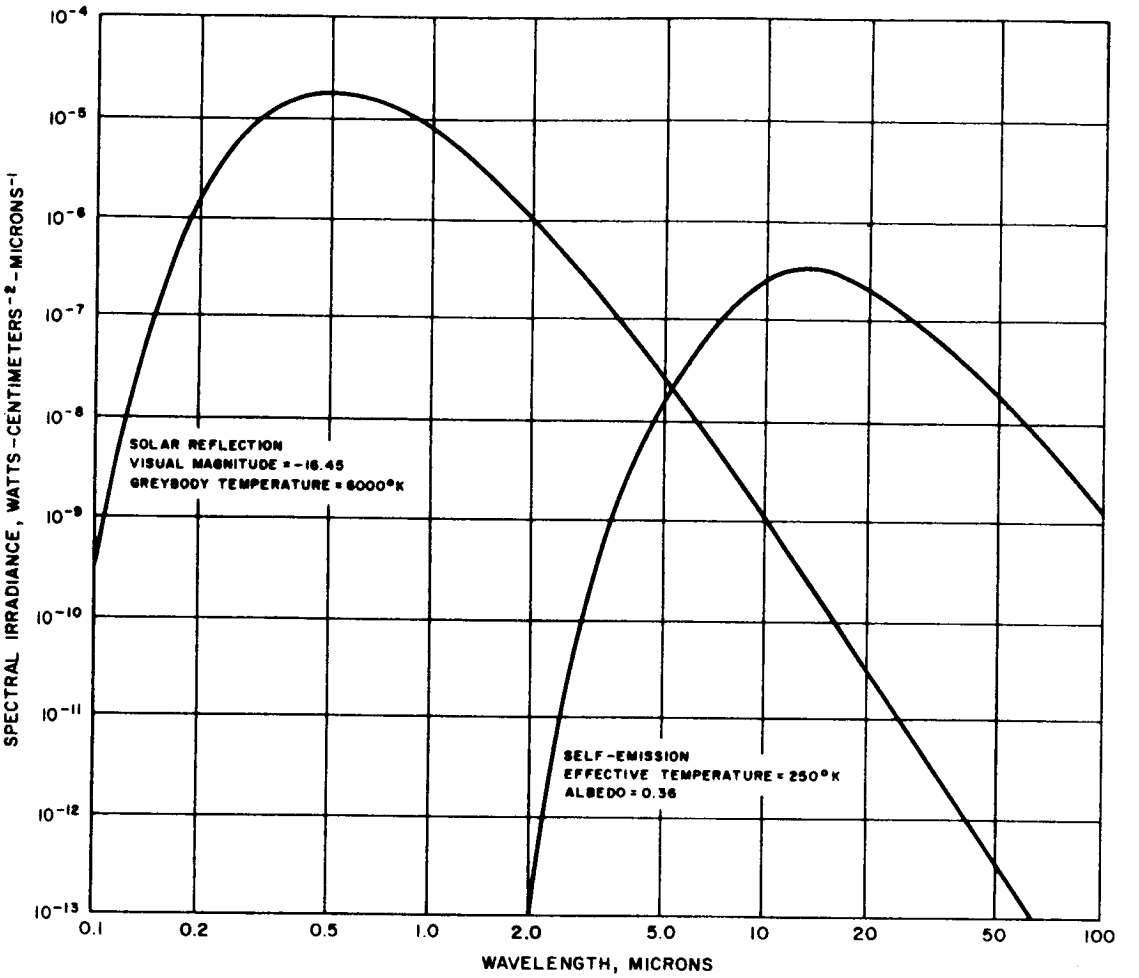
Mars to earth at minimum distance of 78.3(10^6) km



SOURCE: UNIVERSITY OF MICHIGAN, INSTITUTE OF SCIENCE AND TECHNOLOGY. INVESTIGATION OF OPTICAL SPECTRAL REGIONS FOR SPACE COMMUNICATIONS, 2ND INTERIM REPORT, NOV 1962

Figure 8-6. Calculated Maximum Irradiance from Full Moon Outside Earth's Atmosphere

Moon to earth at minimum distance of 384.4(10^3) km



SOURCE: UNIVERSITY OF MICHIGAN, INSTITUTE OF SCIENCE AND TECHNOLOGY. INVESTIGATION OF OPTICAL SPECTRAL REGIONS FOR SPACE COMMUNICATIONS, 2ND INTERIM REPORT, NOV 1962

Figure 8-7. Calculated Maximum Irradiance from Earth at Moon

Earth to moon at minimum distance of 384.4(10^3) km

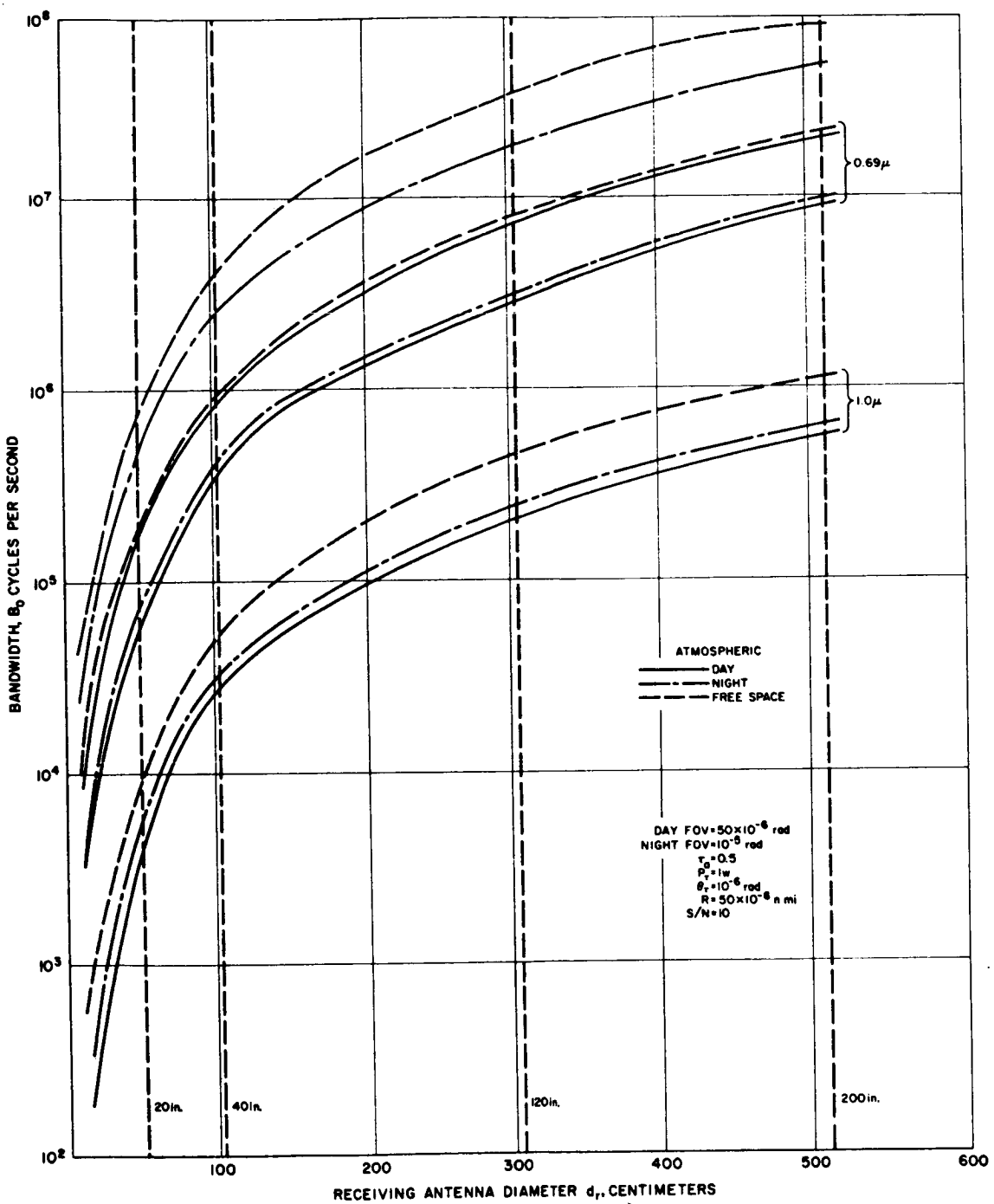


Figure 8-8. Link Comparisons

9. OPERATIONS ANALYSIS

The operations analysis study tasks of the Deep Space Optical Communications Study has been completed and reported on in interim reports. References to this material are given below:

ENERGY SOURCES

First Interim Report, pages 31 – 34.

TRACKING AND POINTING

First Interim Report, pages 35 – 36.
Third Interim Report, pages 89 – 95.

SPACECRAFT STABILIZATION

First Interim Report, pages 36 – 37.

SOURCE-DESTINATION RELATIONSHIPS

First Interim Report, pages 23 – 27.

EARTH-BASED STATIONS

Fourth Interim Report, pages 2 – 13.

DEEP SPACE VEHICLE STATION

Fourth Interim Report, pages 13 – 17.

10. SYSTEMS ANALYSIS

In this section, the material and analyses of the previous sections are integrated to provide a basis for the selection of the "best" system, based on theoretical and practical arguments. In a study of this kind, a compromise between theoretical and practical considerations must continually be made, for the theoretically best system may be well beyond today's practical capability, and today's most practical system may be well below the ultimum system theoretically achievable.

The technique used here has been to select the operating frequency based on practical considerations, which from theoretical considerations implies a certain system. This frequency and detector regime was then analyzed in a system context to decide on the location for the various link elements. The modulation systems which appear to provide the best potential were thus indicated.

Since guidelines of additional study, research, and development are also valuable results of an analysis of this kind, areas in which improvement is desired are also indicated as are the possible effects in technological breakthrough in the overall results of the study.

TRANSMISSION FREQUENCY SELECTION

The approach taken in this study has been to minimize the on-board DSV weight per bit per second of information rate. The main factors contributing to this weight are the size of the entrance optics (antennas) and the weight of the power supply. From the standpoint of the size of the entrance optics the way to go is toward smaller wavelength (higher frequencies), as it is seen from the relationship for the diameter d of a diffraction limited circular aperture:

$$d = 1.22 \frac{\lambda}{\theta} \qquad (10\text{-}1)$$

where θ is the beam cone angle. On the other hand, the maximum information content of an electromagnetic wave within a certain bandwidth B_o, in

presence of uncorrelated (white) additive noise and under the restriction of a signal average power $\bar{P}_s$, diminishes rapidly as the frequency ν is increased beyond $\bar{P}_s/hB_o$. This is clearly illustrated in Figure 7-4. Therefore the frequency $\nu = c/\lambda$ cannot be increased indefinitely without penalty to system performance.

The signal average power $\bar{P}_s$ in Figure 7-4, as well as the additive noise, are referred to the point where the electromagnetic wave will be utilized, the input to the detector. In order to have this power $\bar{P}_s$ in the detector, a power $\bar{P}_L$ is required at the output of the transmitting laser given by the relation (see Equation 8-2)

$$\bar{P}_L = \frac{1}{\tau_r \tau_t \tau_a} \left(\frac{\theta_t R}{d_r} \right)^2 \bar{P}_s \qquad (10\text{-}2)$$

The transmitter and receiver optical transmittances are considered constant for reasons explained in Section 8 while the frequency (wavelength) dependence of the atmospheric transmittance for the standard atmosphere is given in Figure 8-1.

The average input signal power $\bar{P}_s$ for a given constant information rate I is frequency dependent. Figures 10-1 and 10-2 are the same as Figures 7-4 and 7-5 but plotted in a slightly different manner for the present discussion. The selection of a frequency will be based on the assumption that by selecting a proper coding (modulation) scheme, as described in the following section, we can utilize a certain fraction α (say $\alpha = 0.1$) of the "channel capacity," which is the product of the information extraction ability γ of the detector and the information capacity C_w of the electromagnetic wave. Thus, from Figures 10-1 and 10-2 it is clear that, neglecting increased system weight and complexity, larger wavelengths give increased system performance. However, for an earth-space link the transmittance of the atmosphere τ_a must also be taken into account, and low values of τ_a are directly reflected in increased transmitter power required.

That the power required at the detector increases with frequency is evident from the set of curves in Figure 10-2. In order to obtain the required information transmission rate I, the wave capacity should be

$$C_w = \frac{I}{\alpha \gamma(\lambda)} \qquad (10\text{-}3)$$

where α is the constant percent of the channel capacity which a coding scheme can utilize and $\gamma(\lambda)$ is given in Figure 10-1 (in the sequel modified by the quantum efficiency η). Now γ decreases with frequency up to

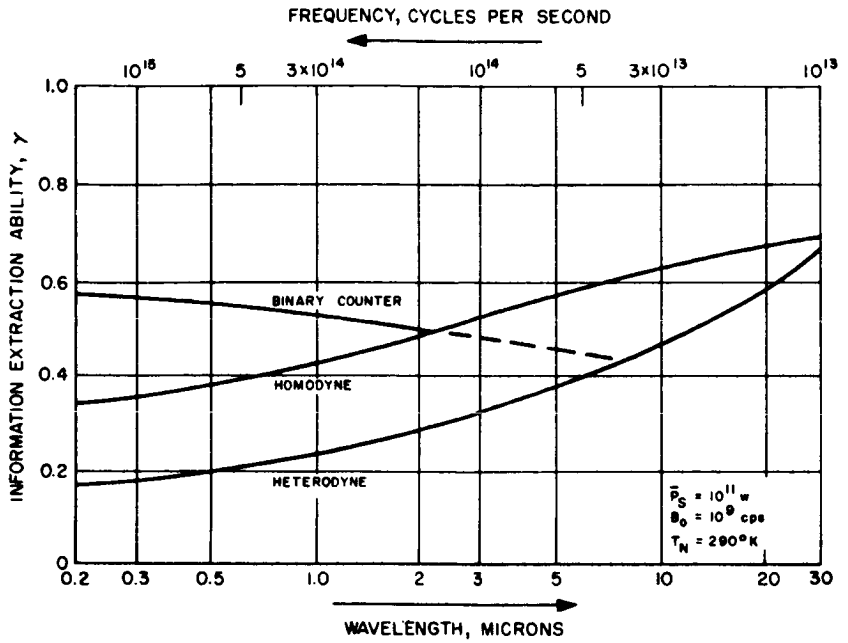


Figure 10-1. Information Extraction Ability

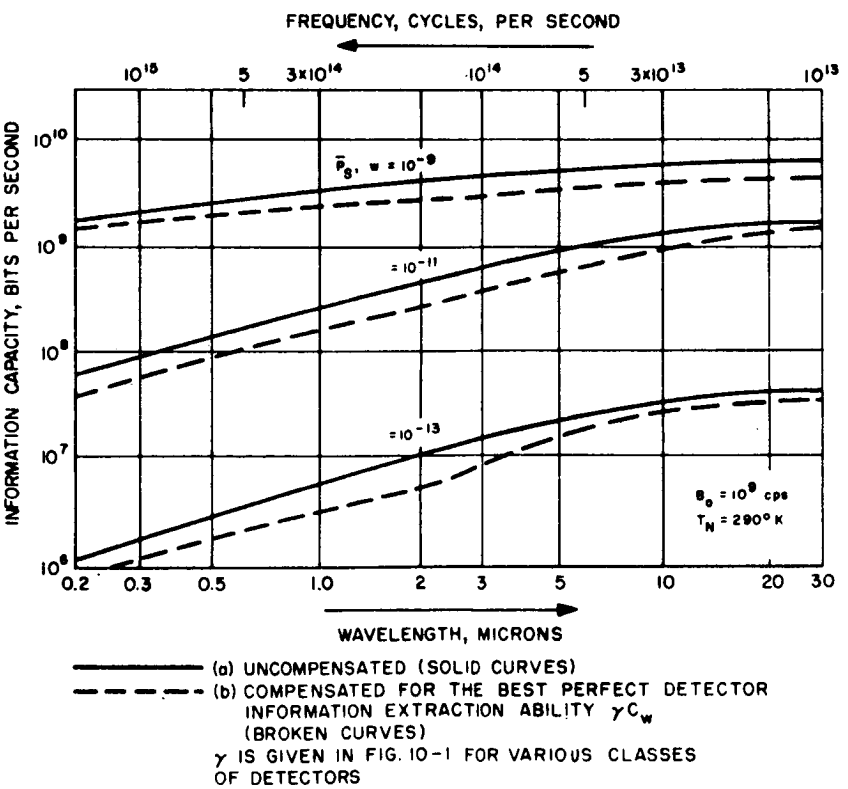


Figure 10-2. Information Capacity in Electromagnetic Wave C_w

frequency 1.3×10^{14} cps (for which homodyne is theoretically the optimum system to use) and from this frequency on, γ increases slowly (when the binary counter is best in the theoretical sense). However, for the wave capacity given by Equation 10-6, the power $\bar{P}_s$ increases with frequency much faster than the reduction in C_w caused by increasing γ. This can be seen easily if Equation 10-6 is written as

$$\gamma C_w = \frac{I}{a} = \text{const} \qquad (10\text{-}4)$$

and the γC_w curves in Figure 10-2 (broken lines) are observed. From laser power considerations, we see from the set of curves in Figure 10-2, plotted for black body background at 290° K and a bandwidth $B_o = 10^9$ cps, it is of interest to choose a frequency in the region above 30 microns. For an earth-based system, however, the atmosphere is opaque in this region, presenting no transmittance window, nor are there any available sources for use as a transmitter.

Figure 10-1 is plotted for the perfect "detector" with efficiency $\eta = 1$. While the quantum efficiency of practical detectors in the region above 1 micron can approach unity as indicated in Figure 10-3, these detectors are severely bandwidth-limited, as discussed in Section 2. For detectors with the bandwidth performance required, the quantum efficiency is below that indicated in the figure. Thus, modification of the curves in Figure 10-1, taking account of the quantum efficiency as a function of bandwidth of the detectors, would be profitable. A more detailed analysis would have to include this functional relationship and determine the optimum band on detector quantum efficiency – bandwidth performance. Finally the power required at the transmitter antenna output may be translated into weight of required power supply and cooling apparatus by taking into account the laser efficiency.

From the discussion above it follows that theoretical considerations favor frequencies in the mid-infrared region of the optical spectrum while practical considerations favor higher frequencies.

The choice of a laser operating in a window in the IR-portion of the spectrum seems highly desirable, based on theoretical considerations. Thus if the high gain of the He-Xe laser at 3.5 microns, for example, could be translated into layer powers, and if its spectral characteristics would permit homodyne operation, it could very well be competitive for future use in a coherent system, provided suitable detectors could be found. Sufficient data is not available to support any contention of this kind, however.

The following considerations will be used to govern the selection of operating frequency:

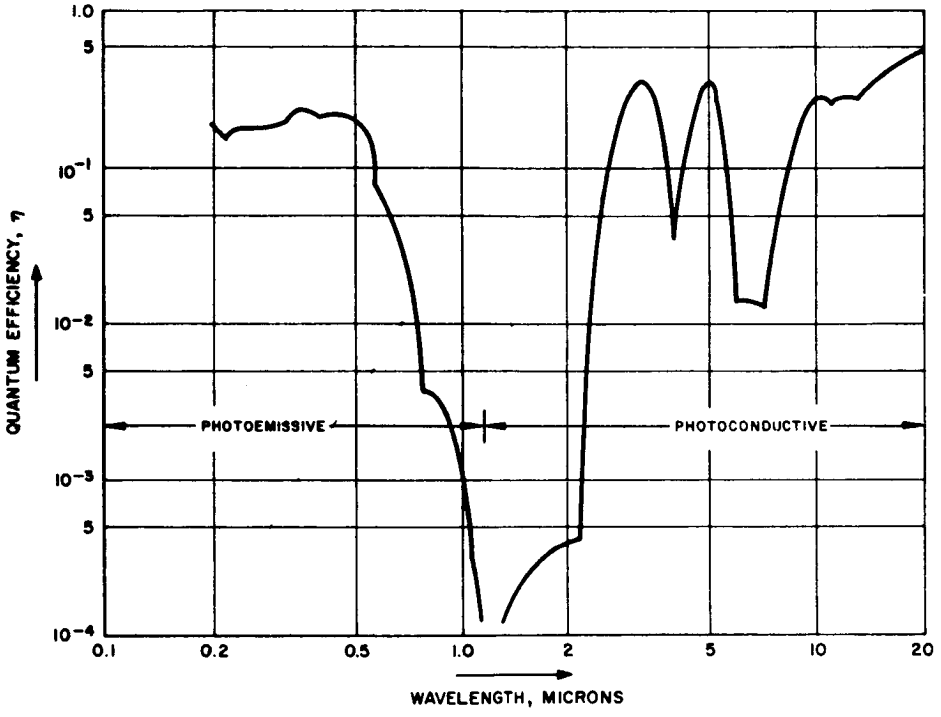


Figure 10-3. Quantum Efficiency of Photodetectors

$$\bar{P}_s = \frac{\bar{P}_t d_r^2 \tau_p \tau_r}{\theta_t^2 R^2} \qquad (10\text{-}5)$$

where $\tau_p \times \tau_i$ is the total transmittance from transmitting antenna to detector (see Equation 10-2).

Assuming that video detection is utilized, it is seen from Equation 7-13 that

$$\frac{S}{N} = \frac{\eta \bar{n}_s^2}{2(\bar{n}_s + \bar{n}_b + \bar{n}_d) B_o} \qquad (10\text{-}6)$$

Assuming that a cooled photodetector is used in which all the dark current photons have been eliminated and $\bar{n}_d \equiv 0$,

$$\frac{S}{N} = \frac{\eta \bar{n}_s^2}{2 B_o (\bar{n}_s + \bar{n}_b)} \qquad (10\text{-}7)$$

Now for an antenna operating in the atmosphere, the principal source of noise is atmospheric and varies with the time of day. Moreover, the receiver field of view required to keep the transmitter under observation as a result of atmospheric image motion, etc., varies. In this analysis we assume 10 micro-radians for night and 50 micro radians for daytime observations, θ_r, for free space links (synchronous satellite or lunar base) is assumed to be $\simeq$ 1 micro-radian due to pointing and tracking limitations. Assuming a background radiance caused by the average noise background power is given by

$$\bar{P}_b = \frac{\pi^2}{16} \Delta\nu \, \theta_r^2 \, d_r^2 \, N_b(\lambda)$$

$$\bar{\mu}_b = \frac{\pi^2}{16} \Delta\nu \, \theta_r^2 \, d_r^2 \, \frac{N_b(\lambda)}{h\nu} \qquad (10\text{-}8)$$

where $N_b(\lambda)$ is background spectral radiance and is related to $H_b(\lambda)$ given in Section 8 by $H_b(\lambda) = N_b(\lambda) \, d\Omega$, where d is the angular field of view of the receiver and is given for small angles by $d = \pi\theta_r^2/4$.

$$\bar{P}_s = \frac{\bar{P}_t d_r^2 \tau_p \tau_r}{\theta_t^2 R^2} \qquad (10\text{-}5)$$

where $\tau_p \times \tau_i$ is the total transmittance from transmitting antenna to detector (see Equation 10-2).

Assuming that video detection is utilized, it is seen from Equation 7-13 that

$$\frac{S}{N} = \frac{\eta \bar{n}_s^2}{2(\bar{n}_s + \bar{n}_b + \bar{n}_d) B_o} \qquad (10\text{-}6)$$

Assuming that a cooled photodetector is used in which all the dark current photons have been eliminated and $\bar{n}_d \equiv 0$,

$$\frac{S}{N} = \frac{\eta \bar{n}_s^2}{2 B_o (\bar{n}_s + \bar{n}_b)} \qquad (10\text{-}7)$$

Now for an antenna operating in the atmosphere, the principal source of noise is atmospheric and varies with the time of day. Moreover, the receiver field of view required to keep the transmitter under observation as a result of atmospheric image motion, etc., varies. In this analysis we assume 10 micro-radians for night and 50 micro radians for daytime observations, θ_r, for free space links (synchronous satellite or lunar base) is assumed to be $\approx$ 1 micro-radian due to pointing and tracking limitations. Assuming a background radiance caused by the average noise background power is given by

$$\bar{P}_b = \frac{\pi^2}{16} \Delta\nu \, \theta_r^2 \, d_r^2 \, N_b(\lambda)$$

$$\bar{n}_b = \frac{\pi^2}{16} \Delta\nu \, \theta_r^2 \, d_r^2 \, \frac{N_b(\lambda)}{h\nu} \qquad (10\text{-}8)$$

where $N_b(\lambda)$ is background spectral radiance and is related to $H_b(\lambda)$ given in Section 8 by $H_b(\lambda) = N_b(\lambda)\, d\Omega$, where $d\Omega$ is the angular field of view of the receiver and is given for small angles by $d\Omega = \pi\theta_r^2/4$.

Solving for B_o in Equation 10-8 gives

$$B_o = \frac{\eta \ \bar{n}_s^2}{2\left(\frac{S}{N}\right)(\bar{n}_s + \bar{n}_b)} \qquad (10\text{-}9)$$

Substituting in Equation 10-9 for n_s from Equation 10-8 gives

$$B_o = \frac{\eta \ \bar{n}_s^2}{2\left(\frac{S}{N}\right)\left(\bar{n}_s + \frac{\pi^2}{16} \Delta\nu \, \theta_r^2 \, d_r^2 \, \frac{N_b(\lambda)}{h\nu}\right)} \qquad (10\text{-}10)$$

The results for $\bar{P}_s$ = 1 watt, λ = 0.5 micron, 0.69 micron, $_{o}$1.0 micron, and for signal-to-noise ratio of 10 with a filter passband = 10 Å is given for the various links in Figure 10-4

It is of interest to note that for systems operating at the same frequency, the size of the ground-based antenna can be said to be "more reasonable" than the corresponding orbiting or lunar-based antenna. For example, at 0.69 micron, a 60-inch telescope on the ground is equivalent to a 40-inch telescope on a synchronous satellite or on the moon, in order to provide a maximum bandwidth of 1 mc. The effect of operating frequency is also quite pronounced, giving larger bandwidths as was expected on less complicated grounds. Note that for λ = 1.0 micron, a 200-inch telescope in free space is required for a 1 mc bandwidth system, while only a 40-inch telescope is required at 0.69 micron. This reflects in part the receiver gain and quantum efficiency of the photodetector dependence on the wavelength involved. The difference in required antenna sizes between free space and atmospheric cases is quite pronounced at significantly lower bandwidths for the large wavelength systems. The difference between day and night operation is relatively insignificant for all but the high frequency case.

The curve strongly supports earth-based operation for incoherent, quantum counter systems with existing detector component technology. Since, neglecting fine structure the transmissivity is approximately constant over the region of the peak of the quantum efficiency of photo-emitter, one should use a laser which operates as close to this peak as possible. It must be noted that this selection of an earth-based receiver system is contingent upon the use of incoherent reception only, and the advent of coherent receivers in the IR portion of the spectrum could possibly favor a free-space receiver system with performance which is superior.

MODULATION METHOD SELECTION

The previous analysis has narrowed the selection of a modulation method for a deep space communication system to a choice of one of the

incoherent reception techniques. The comparative advantages and disadvantages of the incoherent forms of reception have been presented in the section on modulation systems. The results of that analysis are presented below.

1) Incoherent analog modulation techniques may be eliminated from consideration because of their lower communication efficiencies compared to the incoherent, time-quantized forms of modulation.

2) PPM and PCM are the most efficient forms of time-quantized modulation.

3) For a high background noise environment, in the absence of "signal" noise, PCM is superior to PPM. In general, narrow optical filters or heterodyning techniques are necessary to take advantage of the lower bandwidth of PCM compared to PPM for the same information rate.

4) For a low amount of background noise, PPM may take precedence over PCM if bandwidth occupancy is not a consideration and power conservation is the primary design factor.

5) Incoherent pulse code modulation transmission takes three main forms: PCM/AM, PCM/FSK, and PCM/PL. They exhibit nearly the same communication system efficiency.

6) With a laser oscillator of peak power limitation, PCM/AM will be restricted to an operation at one-half the average power transmission of PCM/FSK and PCM/PL.

The theoretical selection of an optimum incoherent modulation method cannot be made among the resultant choices of PPM, PCM/FSK, and PCM/PL until a much deeper understanding is available of the system noise environment and detection processes. However, since the three systems are expected to lie in the same general efficiency range, the choice between them for a present or near future system can be made on the basis of implementation requirements. Toward this goal, the following capsule review of the implementation requirements of the systems is presented.

Continuous wave gaseous lasers suitable for the generation of PCM carrier waveforms presently exist. They exhibit an extremely high degree of spectral purity and coherency, and are capable of being collimated to small beam angles. Their weights are reasonable for spacecraft use, but gas lasers tend to be relatively long devices. The average power available with a gas laser is relatively low for deep space applications. However, it is comparable to the CW laser injection diode and the $Dy^{2+}:CaF_2$ and $Nd^{3+}: CaWO_4$ solid state laser which are the only other devices that could be used for PCM transmission.

Many physical problems presently exist with the CW semiconductor diode laser. It is difficult to reliably induce laser action in an injection diode. Also, the spectral purity and frequency drift are far below the

capabilities of the gas laser. The laser diode beam angle is much broader than that of a gas laser, and collimation of the beam presents problems because of its fan shape. A further problem exists in that temperature-stable cryogenic cooling is required for laser diodes. On the bright side, the semiconductor laser is a small, lightweight device which may be mounted within the system optics rather easily. The problems associated with the use of laser diodes do not appear unsolvable, but for an immediate system it does not appear feasible to use these devices for PCM transmission.

The solid state laser suffers most of the drawbacks of the semiconductor laser including poor spectral purity and the requirement of cryogenic cooling. Its continuous carrier waveform is also subject to a considerable amount of spiking.

For PCM transmission with a gas laser, no practical methods of internal modulation exist for either frequency shift keying or polarization modulation. Polarization modulation can be performed by a cascaded electro-optic modulator or a Fabry-Perot modulator. Frequency modulation, however, cannot be performed with reasonable efficiency or quality with any present modulator.

PPM is most easily handled by the semiconductor diode operating in a pulse mode. However, the pulsed laser diode is subject to the same operational limitations as its continuous wave counterpart.

Modulation of a laser diode for PPM is performed internally by current pumping. However, the high current, modulation pulses required by the diode must be generated outside the diode by modulator switching circuits. For high information rates, the only switching component that is capable of handling the fast switching speeds is the avalanche transistor. Unfortunately, this device presently limits the full utilization of the laser diode power output because of its own internal power dissipation.

Based upon these considerations of laser and modulator availability and capability, the communication system that seems most feasible to implement at present is the PCM polarization modulation system using a CW gas laser.

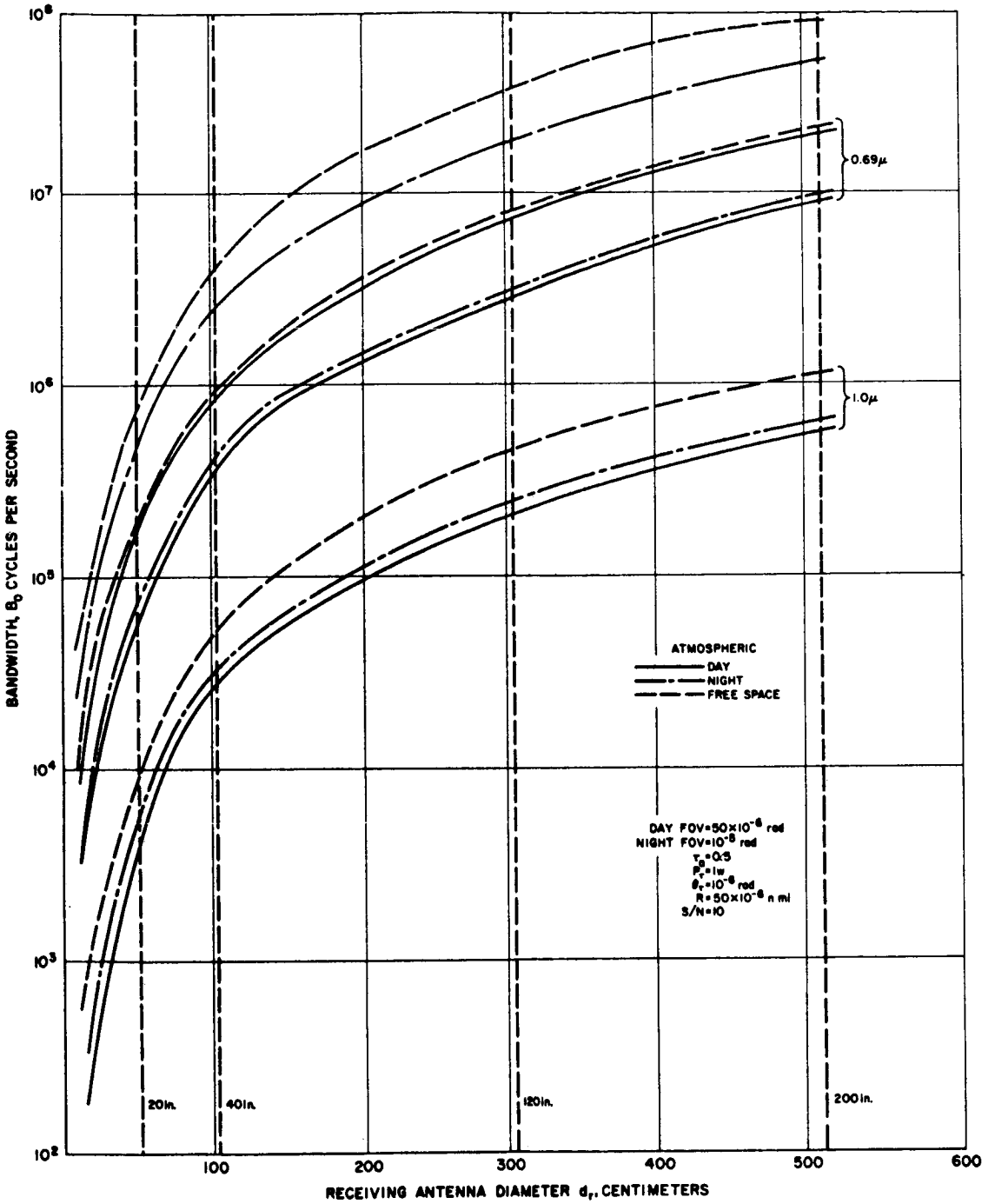


Figure 10-4. Link Comparisons

11. PRELIMINARY SYSTEMS DESIGN

The systems analysis study of the previous section has reduced the search for an optimum deep space communication system for the near future to two systems: pulse position modulation, using a gallium arsenide diode, and PCM polarization modulation, using a CW gas laser.

For an immediate communication system the practical problems of implementing a PPM system have eliminated it in favor of the PCM system which can be reliably implemented with present system components. For this reason the following discussion has been limited to a preliminary systems design analysis of the PCM polarization modulation communication system.

The fourth interim report of the Deep Space Optical Communication Study (Reference 11-1) presented a preliminary systems design of the same two systems on a more futuristic basis. The philosophy of that analysis was to determine the laser output power and system bandwidth required for a given information rate that was tailored for the transmission of real time television. The high information rate requirements led to system power and bandwidth specifications that are above the present projected state of the art near-term systems.

For this preliminary system design analysis a more conservative approach has been taken to give an indication of the configuration and performance characteristics of the PCM/PL system as it might be implemented in the very near future. The philosophy of this design analysis has been to set capability limitations on the optical components of the system and the associated electronic equipment in terms of both speed and power. The result of the analysis is a preliminary configuration of the system and an evaluation of its performance characteristics as a function of information sample rate and laser output power.

PCM POLARIZATION MODULATION SYSTEM

In PCM polarization modulation transmission, the state of polarization of the transmitted beam is set in one of two possible states to represent binary ones or zeros of the message. Typically, the two are orthogonal linear polarizations or right-hand and left-hand circular polarization. The state of

polarization at the receiver will be determined by separating the two significant polarizations by means of a polarizing prism. Light from the daytime sky is partially unpolarized and partially linearly polarized and, therefore, it might be detected unequally by the two channels if linear polarization were transmitted. For this reason it is desirable to transmit circularly polarized light.

A typical PCM polarization modulation communication system is shown in Figure 11-1. A CW helium-neon gas laser operating at 6328 Å has its linearly polarized output converted into either a right or left circularly polarized light by a cascaded element electro-optical modulator. The output of the modulator is collimated by a reflecting telescope. The receiver telescope collects the transmitted light signal and passes it through a passive optical filter of bandwidth 10 Å or less. The polarization resolving optics utilize a quarter-wave plate to transform the two opposite circular polarizations into orthogonal linear polarizations. The resulting beam is spatially resolved and focused on separate detectors. After detection the right and left channel signals are subtracted and the resultant is applied to a threshold electrical polarity detector.

The implementation problems associated with this type of PCM polarization modulation system are due to power limitations of the gas laser and speed and power limitations of the modulator circuits. The present gas lasers operate to about 100 mw of average power with a 1 watt upper limit predicted for the near future. Electro-optic modulators can operate up to a rate of 50 mc with reasonable efficiency.

As mentioned in the systems analysis discussion, the optimum operating wavelength point for an incoherent communication system such as the PCM/PL system will be at the peak of the photodetector quantum efficiency curie (Figure 10-3, Section 10). The gas laser operating nearest the quantum efficiency peak is the helium neon gas laser operating at a wavelength of 0.63 micron.

PHYSICAL DESIGN DETAILS

The physical design details of the construction of the optical and mechanical portions of the communication system have been presented in the fourth interim report of the Deep Space Optical Communication Study. The pertinent references are given below.

1) Mounting to the Spacecraft

   Fourth Interim Report, pages 55-56.

2) Primary Optical System

   Fourth Interim Report, page 57.

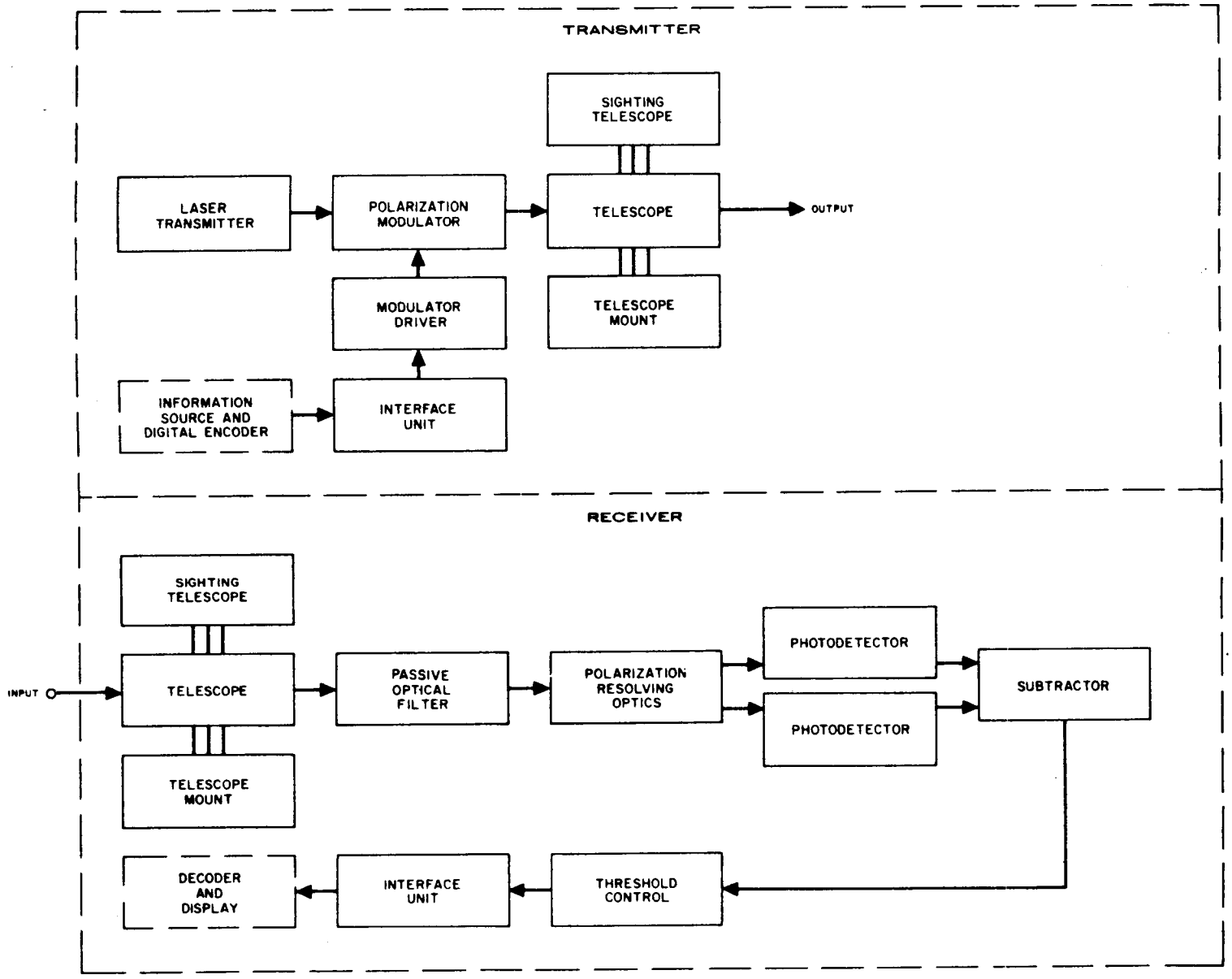


Figure 11-1. PCM Polarization Modulation Communication System

3) Isolation Gimbal

Fourth Interim Report, pages 57-60.

4) Divisional Optical Systems

Fourth Interim Report, pages 61-63.

PERFORMANCE EVALUATION

The performance of the PCM system may be evaluated by the use of the following equations* relating the signal quanta per sampling period, S, and the noise quanta per sampling period, N, to the system parameters. The definition and example values of the parameters are given in Table 11-1.

$$N = \frac{Q_B(\lambda)\,\Delta\lambda\, d_R^2\, \theta_R^2\, \pi^2\, T_A\, T_R \eta \tau}{16} \tag{11-1}$$

$$S = \frac{d_R^2\, T_T\, T_A\, T_R\, T_S\, \eta \tau}{\theta_T^2\, R^2 (hc/\lambda)}\, P_T \tag{11-2}$$

The relationship between Equations 11-1 and 11-2 (References 11-2 and 11-3) is given in Figure 11-2 for the PCM form of modulation.

The evaluation procedure has been to calculate N as a function of τ and S as a function of τ and P_T. The relationships are shown below for the PCM system operating at a wavelength of 0.63 micron for the three cases of background conditions.

Daytime sky

$$N = 1.65 \times 10^8\, \tau \tag{11-3}$$

$$S = 2.27 \times 10^6\, \tau\, P_T \tag{11-4}$$

Nighttime Mars

$$N = 3.9 \times 10^6\, \tau \tag{11-5}$$

$$S = 2.27 \times 10^6\, \tau\, P_T \tag{11-6}$$

---

*These equations are developed in Section 8.

TABLE 11-1. PARAMETER VALUES FOR A DSV-EARTH COMMUNICATION LINK

Parameter	Symbol	Value
Distance	R	9.3×10^{12} cm
Transmitter beam diameter	θ_T	5×10^{-6} radians
Receiver field diameter	θ_R	
Daytime		10^{-4} radians
Nighttime		2.5×10^{-5} radians
Receiver aperture diameter	d_r	300 cm
Receiver optical filter noise bandwidth	$\Delta\lambda$	10^{-3} microns
Transmitter optical transmittance	T_T	0.85
Atmospheric optical transmittance	T_A	0.75
Receiver optical transmittance	T_R	0.60
Atmospheric scintillation factor	T_S	0.90
Receiver phototube quantum efficiency	η ($\lambda_1 = 0.63$ micron)	0.05
	($\lambda_2 = 0.84$ micron)	0.003
Spectral photon radiance	$Q_B(\lambda)$	
Daytime sky	($\lambda_1 = 0.63$ micron)	10^{16} photon / sec cm^2 srμ
	($\lambda_2 = 0.84$ micron)	5×10^{15} photon / sec cm^2 srμ
Nighttime Mars	($\lambda_1 = 0.63$ micron)	5×10^{15} photon / sec cm^2 srμ
	($\lambda_2 = 0.84$ micron)	3×10^{15} photon / sec cm^2 srμ
Nighttime sky		$\cong 0$
Sample error rate	P_e'	10^{-2}

Nighttime sky

$$N = 0 \tag{11-7}$$

$$S = 2.27 \times 10^{6} \tau P_T \tag{11-8}$$

The results of the performance analysis for the PCM/PL systems are shown in Figure 11-3. In the figure the information sample period is plotted as a function of required transmitter power. The large power requirements for the daytime transmission case are due in part to the fact that the filter bandwidth of 10^{-3} micron is much greater than the transmission bandwidth. If a narrow-band filter could be employed through the use of a laser preamplifier active filter or the application of superheterodyning techniques, the power requirements could be reduced considerably for this case.

REFERENCES

11-1. "Deep Space Optical Communication Study," Fourth Interim Report, pages 55-72.

11-2. Ibid, page 36.

11-3. Ibid, page 47.

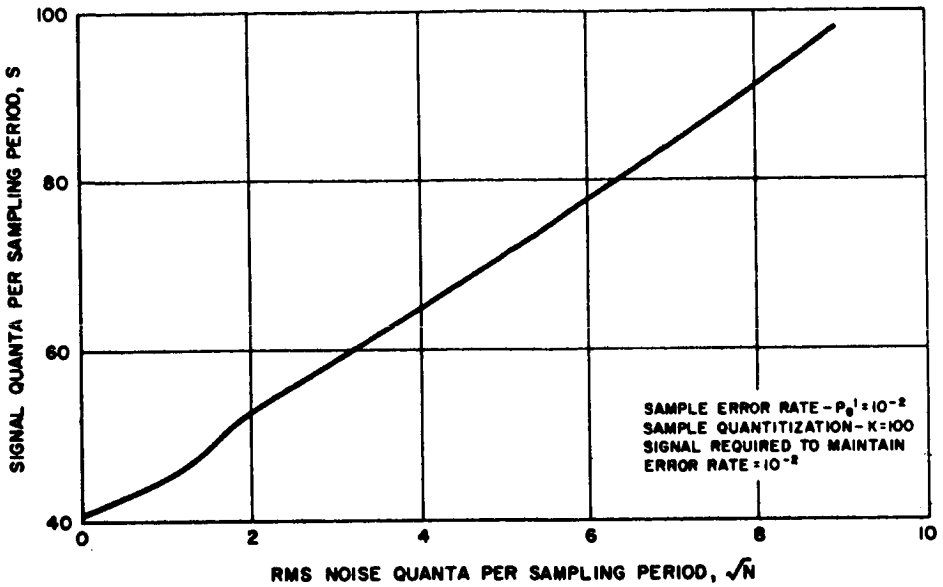


Figure 11-2. Pulse Code Modulation Signal-to-Noise Relationship

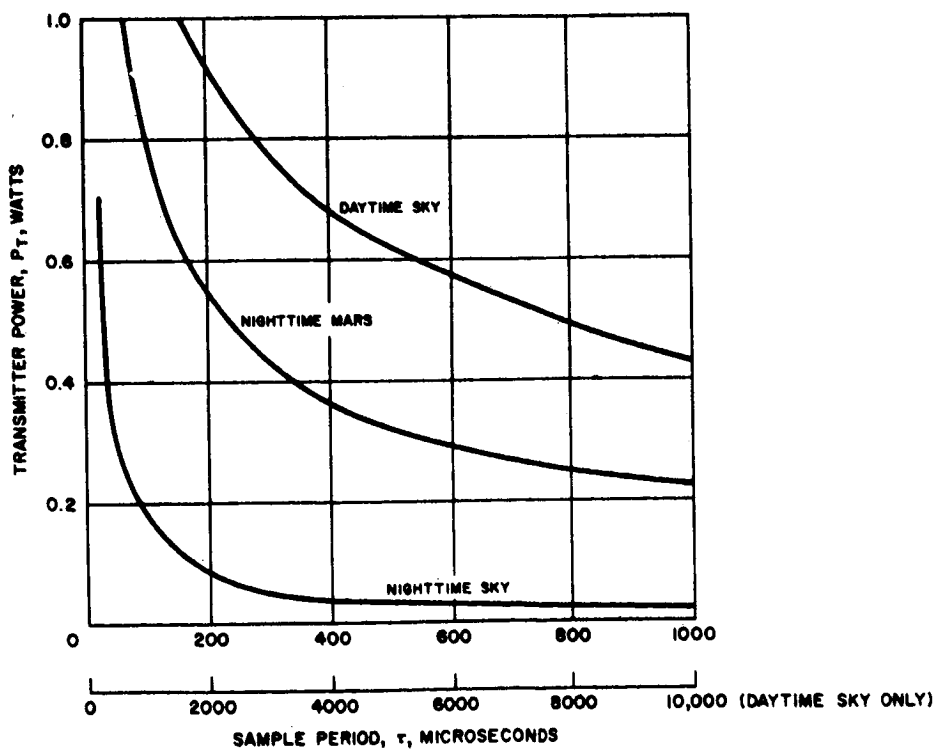


Figure 11-3. Pulse Code Modulation Sample Information Rate Versus Transmitted Power

12. CONCLUSIONS AND RECOMMENDATIONS

The results of this preliminary systems analysis indicate that optical communication from deep space is feasible and competitive with microwave communication. While the areas of component development are rapidly expanding, the results of this study favor the use of an earth-based receiver system utilizing incoherent (quantum-counter) detection at frequencies in the visible portion of the spectrum.

Additional study is required of coherent reception to place it in the proper context with respect to incoherent reception, but theoretical studies indicate that potential exists for improved system performance utilizing coherent reception in the IR portion of the spectrum, and possibly from a synchronous satellite with a microwave or optical link to earth. A definite decision can be reached only through increased experimental knowledge of atmospheric effects on coherent reception and improved detection performance at IR.

The currently competitive systems appear to be the following:

1) Earth-based

   CW laser utilizing PCM-polarization modulation
   Pulsed laser utilizing PPM

2) Earth- or satellite-based

   Coherent reception

It is recommended that the following be initiated:

1) An advanced system study to compare the relative merits of these three systems, based on a deeper analysis of their capabilities and on experimental results which could be obtained by building one of them.

2) The fabrication of an experimental CW PCM-polarization modulation system to provide the experimental test-bed indicated

above, as well as to establish the limitations of this competitive system. This system was chosen because it appears to be the one which can be most easily implemented at the present time.